

UNIVERSITA' VITA-SALUTE SAN RAFFAELE
CORSO DI DOTTORATO DI RICERCA INTERNAZIONALE
IN MEDICINA MOLECOLARE
CURRICULUM IN NEUROSCIENZE E NEUROLOGIA
SPERIMENTALE

**A MOLECULAR AND
ENVIRONMENTAL APPROACH TO
DISSECT DISEASE HETEROGENEITY
IN MULTIPLE SCLEROSIS**

DoS: Dr. Federica Esposito

Federica Esposito

Second Supervisor: Prof. Maja Jagodic

Tesi di DOTTORATO di RICERCA di Antonino Giordano

Matr. 015567

Ciclo di dottorato: XXXV

SSD: MED/26

Anno Accademico 2021/2022

CONSULTAZIONE TESI DI DOTTORATO DI RICERCA

Il/la sottoscritto/I Antonino Giordano
Matricola / registration number 015567
Nat_ a/ born at Cefalù (PA)
il/on 28/09/1989

autore della tesi di Dottorato di ricerca dal titolo / *author of the PhD Thesis titled*

A MOLECULAR AND ENVIRONMENTAL APPROACH TO DISSECT DISEASE HETEROEogeneity IN MULTIPLE SCLEROSIS

AUTORIZZA la Consultazione della tesi / *AUTHORIZES the public release of the thesis*

NON AUTORIZZA la Consultazione della tesi per 12 mesi / *DOES NOT AUTHORIZE the public release of the thesis for 12 months*

a partire dalla data di conseguimento del titolo e precisamente / *from the PhD thesis date, specifically*

Dal / *from* 28/03/2023 Al / *to* 27/03/2024

Poiché / *because*:

l'intera ricerca o parti di essa sono potenzialmente soggette a brevettabilità/ *The whole project or part of it might be subject to patentability;*

ci sono parti di tesi che sono già state sottoposte a un editore o sono in attesa di pubblicazione/ *Parts of the thesis have been or are being submitted to a publisher or are in press;*

la tesi è finanziata da enti esterni che vantano dei diritti su di esse e sulla loro pubblicazione/ *the thesis project is financed by external bodies that have rights over it and on its publication.*

E' fatto divieto di riprodurre, in tutto o in parte, quanto in essa contenuto / *Copyright the contents of the thesis in whole or in part is forbidden*

Data /Date 23/02/2023 Firma /Signature 

DECLARATION

This thesis has been:

- composed by myself and has not been used in any previous application for a degree. Throughout the text I use both 'I' and 'We' interchangeably.
- has been written according to the editing guidelines approved by the University.

Permission to use images and other material covered by copyright has been sought and obtained.

All the results presented here were obtained by myself, except for:

- 1) Imputation of the genetic data was performed by Dr. Ferdinando Clarelli (Laboratory of Human Genetics of Neurological Disorders, Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy);*
- 2) The wet experiments for methylation profiling with Illumina EPIC arrays described were performed by Dr. Elisabetta Mascia (Laboratory of Human Genetics of Neurological Disorders, Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy);*
- 3) Vitamin D measurement was performed at the Centre Hospitalier Universitaire de Toulouse (France) and supervised by Dr. Béatrice Pignolet and prof. Roland Liblau;*
- 4) The electronic version of the environmental questionnaire was elaborated by Dr. Misra Kaalindi (Laboratory of Human Genetics of Neurological Disorders, Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy);*
- 5) Plasma and CSF neurofilament measurement was performed at the Neuroimmunology Unit, Centre for Molecular Medicine, Karolinska Institute, Stockholm, Sweden;*
- 6) Acquisition and processing of brain MRI and quantification of the paramagnetic rim lesions was performed by Prof. Mara Rocca and Dr. Paolo Preziosa at the Neuroimaging of CNS White Matter Unit, Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy.*

All sources of information are acknowledged by means of reference.

ACKNOWLEDGEMENTS

I want to acknowledge all my colleagues working at the Laboratory of Human Genetics of Neurological Disorders at IRCCS San Raffaele Hospital, headed by Dr. Federica Esposito, without whom this work would not have been possible.

We also acknowledge all the staff of the Multiple Sclerosis Centre and the Neurological Department of IRCCS San Raffaele Hospital (Milan, Italy), headed by Prof. Massimo Filippi, for the fundamental contribution to the enrollment and blood sampling of patients.

A special thanks goes to our partners, the Centre Hospitalier Universitaire de Toulouse (France) coordinated by prof. Roland Liblau and Dr. Béatrice Pignolet, and the Neuroimmunology Unit of the Center for Molecular Medicine at the Karolinska Institutet (Stockholm, Sweden), coordinated by Prof. Maja Jagodic and Prof. Ingrid Kockum. An additional acknowledgement goes to Prof. Lars Alfredsson and Prof. Jan Hillert from the Karolinska Institutet, that were responsible for the Swedish MS Registry.

Part of the present work has been done within the FindingMS project (<https://www.findingms.com/>), which is supported by a grant from the ERA-Net Cofund Joint Transnational Calls ERA PerMed.

We also acknowledge the MultipleMS consortium (<https://www.multiplems.eu>) for supporting part of the experiments described in Chapter 7. The MultipleMS consortium is funded by an EU-Horizon 2020 Grant (no. 733161).

Our last and biggest acknowledgement goes to all the people with MS that were part of the studies described in this work, for their willing of contributing to research advancements.

“Diseases do not read textbooks.”

ABSTRACT

Multiple Sclerosis (MS) is very heterogenous in its clinical course. Over the years, some patients develop no or mild disability, while in others the disease is extremely active and leads to a rapid deterioration of neurological functions that severely affects their quality of life. Despite great advancements in understanding some of the mechanisms related to the disease and the availability of many drugs that can hamper its manifestations, the biological basis underlying such heterogeneity is still largely unknown.

In this PhD project, we aimed to study some of the aspects that characterize the clinical diversity of MS, adopting measures of disease activity, severity, and progression. The ultimate goal is to pave the way to treatment tailoring and potentially to future drug development. To achieve our aims, we took advantage of a broad set of information on the clinical, molecular (genetic, epigenetic, transcriptomic, neurofilaments), and epidemiological (environmental factors) level, to gain a comprehensive insight on some of the explored mechanisms.

In the genetic study of disease activity, we were able to prioritize some loci that are fundamental in the modulation of the immune system, as for the Semaphorins and the NAMPT, and that represent promising candidates for future targeted validation. In this regard, we also provide evidence that vitamin D levels causally affect disease activity.

As a key point, in this project multiple independent evidence coming from the genetic, the epigenetic and the environmental studies pointed out a role of genes related to the activity of sex hormones, strongly suggesting that their known effect on the immune system is responsible for disease manifestations, and that puberty is the fundamental scenario in which the MS pathobiology takes place.

Finally, we discovered a novel locus in the *HIF1A* gene that affects the susceptibility to develop secondary progressive MS, likely driving chronic silent inflammation through the response to oxidative stress, as demonstrated by multiple levels of evidence involving gene expression, methylation, pharmacogenomics, neurofilament levels and the paramagnetic rim lesions.

In conclusions, our findings provide a meaningful insight on different processes that are important in driving clinical heterogeneity in MS, prompting the effectiveness of an integrated approach when studying complex multifactorial diseases.

TABLE OF CONTENTS

List of acronyms and abbreviations -----	1
List of figures and tables -----	2
1. Introduction: Multiple sclerosis -----	7
<i>1.1 Epidemiology and clinical presentation</i> -----	7
<i>1.2 Diagnosis</i> -----	9
<i>1.3 Pathogenesis</i> -----	9
1.3.1 Genetic factors-----	11
1.3.2 Environmental factors-----	11
1.3.2.1 Epstein-Barr Virus-----	12
1.3.2.2 Cigarette smoking-----	12
1.3.2.3 Vitamin D-----	13
1.3.2.4 Gut microbiota, obesity, and other factors-----	13
<i>1.4 Treatment</i> -----	14
<i>1.5 Disease activity in MS: key elements</i> -----	16
2. Aims of the work -----	20
RESULTS -----	21
3. Study 1: Genetic factors underlying disease activity -----	21
3.1 Results-----	22
3.1.1 Meta-analysis results-----	22
3.1.2 Gene Ontology enrichment and pathway analysis-----	24
3.2 Considerations-----	28
4. Study 2: Epigenetic factors underlying disease activity -----	31
4.1 Results-----	32
4.1.1 Identification of Differentially Methylated Positions (DMP)-----	32
4.1.2 Identification of Differentially Methylated Regions (DMR)-----	34
4.1.3 Methylation and expression Quantitative-Trait-Loci effect in AMH-----	36
4.1.4 Sensitivity analysis-----	38
4.2 Considerations-----	38
5. Study 3: Vitamin D and disease activity -----	42
5.1 Results-----	43
5.1.2 Genetic variation in vitamin D affects disease activity-----	43
5.1.3 Mendelian Randomization analysis-----	46
5.2 Considerations-----	46
6. Study 4: Environmental factors and disease severity -----	50
6.1 Results-----	51
6.1.1 Factors provoking changes in sex hormones-----	51
6.1.3 Alcohol consumption-----	54

6.1.3.5 Alcohol consumption in men-----	56
6.2 Considerations-----	58
6.2.1 Sex hormones and disease severity -----	58
6.2.2 Body weight and disease severity -----	58
6.2.3 Alcohol consumption and disease severity -----	59
6.2.4 General considerations-----	60
7. Study 5: Genetic factors driving chronic silent inflammation -----	61
7.1 Results -----	62
7.1.1 Genetic study on iron metabolism prioritizes a locus HIF1A -----	62
7.1.2 eQTL and mQTL in MS patients-----	65
7.1.3 Impact of the HIF1A variant on NFL levels-----	66
7.1.4 HIF1A variant and response to Disease-Modifying Treatment (DMT)-----	67
7.1.5 The genetic variant in HIF1A affects the PRL burden -----	69
7.2 Considerations-----	71
8. Discussion-----	75
9. Methods -----	79
9.1 Study 1: Genetic factors underlying disease activity-----	79
9.1.1 Inclusion criteria and NEDA-3 definition-----	79
9.1.2 Genotype imputation and QC in the OSR cohort -----	80
9.1.3 Genotype imputation and QC in the CHUT cohort-----	82
9.1.4 Genome-Wide Association Study (GWAS) and meta-analysis-----	82
9.1.5 Variant to genes -----	83
9.1.6 Gene ontology enrichment and pathway analysis-----	83
9.2 Study 2: Epigenetic factors underlying disease activity -----	83
9.2.1 Study cohort-----	83
9.2.2. PBMC extraction and methylation profiling-----	84
9.2.3 Quality Controls (QC) -----	84
9.2.4 Differential methylation analysis-----	85
9.2.5 Gene ontology enrichment analysis-----	85
9.2.6 MQTL and eQTL effect-----	85
9.3 Study 3: Vitamin D and disease activity -----	87
9.3.1 Study cohort-----	87
9.3.2 Vitamin D measurement and analysis -----	87
9.3.3 Mendelian Randomization analysis-----	88
9.4 Study 4: Environmental factors and disease severity -----	90
9.4.1 Study cohort-----	90
9.4.2 Environmental questionnaire -----	90
9.4.3 Statistical analysis-----	90
9.4.4 Definition of categories of alcohol consumption -----	91
9.5 Study 5: Genetic factors driving silent inflammation in progressive MS-----	91
9.5.1 Study cohorts for the genetic analysis-----	91
9.5.2 Genome-wide quality controls in the OSR cohort -----	92
9.5.3 Selection of SNPs in iron metabolism genes -----	92
9.5.4 Genetic association analysis in the discovery cohort-----	93
9.5.5 Replication of the genetic association -----	93
9.5.6 RNA-seq and eQTL effect in PBMC -----	94
9.5.7 Plasma NFL -----	94
9.5.8 Cerebrospinal fluid NFL -----	95
9.5.9 Pharmacogenomic study of DMF -----	95
9.5.10 Paramagnetic Rim Lesions (PRL) analysis-----	95

9.5.10.1 MRI acquisition-----	95
9.5.10.2 Conventional MRI analysis -----	96
9.5.10.3 SWI processing -----	96
9.5.10.4 Quantification of paramagnetic rim lesion number and volume -----	97
9.5.10.5 PRL volume analysis-----	97
References -----	99
APPENDIX-----	123
<i>Supplementary Tables</i> -----	<i>123</i>
Table S3.1 & Table S3.1 -----	123
Table S7.1 & Table S7.2 -----	125
<i>Supplementary Figures</i> -----	<i>137</i>
Figure S2.1: Overview of the work-----	137
Figure S6.1: Environmental questionnaire -----	138

List of acronyms and abbreviations

AMH	Anti-Mullerian Hormone
ARMSS	Age-Related Multiple Sclerosis Severity
ARR	Annualized Relapse Rate
BL	Baseline
CHUT	Centre Hospitalier Universitaire de Toulouse
CNS	Central Nervous System
CpG	Cytosine-phosphate-guanine dinucleotide
CSF	Cerebrospinal Fluid
DC	Dendritic Cells
DMF	Dimethyl Fumarate
DMP	Differentially Methylation Positions
DMR	Differentially Methylated Regions
EAE	Experimental Autoimmune Encephalomyelitis
EBV	Epstein-Barr Virus
EDA	Evidence of Disease Activity
EDSS	Expanded Disability Status Scale
EF	Environmental Factors
EQ	Environmental Questionnaire
eQTL	Expression Quantitative Trait Loci
FDR	False Discovery Rate
GO	Gene Ontology
GWAS	Genome-Wide Association Study
HRC	Haplotype Reference Consortium
IBD	Identity By Descent
IMSGC	International Multiple Sclerosis Genetics Consortium
IWV	Inverse Variance Weighted
Kb	Kilobase
KEGG	Kyoto Encyclopedia of Genes and Genome
LD	Linkage Disequilibrium

MAF	Minor Allele Frequency
mQTL	Methylation Quantitative Trait Loci
MR	Mendelian Randomization
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
MSFC	Multiple Sclerosis Functional Composite
MSSS	Multiple Sclerosis Severity Score
NEDA	No Evidence of Disease Activity
NFL	Neurofilament light chain
OR	Odds Ratio
OSR	IRCCS San Raffaele Hospital
PBMC	Peripheral Blood Mononuclear Cells
PCA	Principal Component Analysis
PIRA	Progression Independent of Relapse Activity
pNFL	Plasma NFL
PRL	Paramagnetic Rim Lesions
QC	Quality Controls
QQ	Quantile-Quantile
RR	Relapsing-Remitting
SNP	Single Nucleotide Polymorphism
SP	Secondary Progressive

List of figures and tables

Figures	
Figure 1.1	Prevalence of MS in the world.
Figure 1.2	Mechanisms of myelin damage in multiple sclerosis.
Figure 1.3	Paramagnetic Rim Lesions, adapted from Martire et al, 2022.
Figure 3.1	QQ-plot showing the genomic inflation factor (λ) for the meta-analysis.

Figure 3.2	Manhattan plot of the genome-wide meta-analysis between OSR and CHUT.
Figure 3.3	Volcano plot showing the results of the Gene Ontology enrichment analysis.
Figure 4.1	Schematic representation of Differentially Methylated Positions (DMP) and Differentially Methylated Regions (DMR).
Figure 4.2	The identified Differentially Methylated Region in the AMH gene.
Figure 4.3	Effect of the rs2240656 variant on methylation and gene expression in MS.
Figure 5.1	Box-dot plot showing the association of oral contraceptive use, menopause, and pregnancy with the ARMSS score.
Figure 5.2	Impact of BMI at the time of diagnosis on future disease severity.
Figure 5.3	Alcohol consumption and MS severity in women.
Figure 5.4	Alcohol consumption and disease severity in men.
Figure 6.1	Disease activity at 2 years is affected by baseline vitamin D levels.
Figure 6.2	Circular Manhattan plot showing the association with NEDA-3 status for the SNPs with known role on vitamin D level variation in controls (Revez et al, 2020)
Figure 6.3	Mendelian Randomization analysis of vitamin D and NEDA-3 status.
Figure 7.1	Regional plot for the top associated signal in chromosome 14 in the OSR cohort.
Figure 7.2	Expression Quantitative-Trait-Loci (eQTL) effect in PBMC from naive RR-MS patients.
Figure 7.3	Effect of the rs11621525 variant of plasma and CSF NFL levels in RR-MS.
Figure 7.4	Impact of the rs11621525 variant on the NFL levels after DMT.
Figure 7.5	Effect of the variant in HIF1A on the clinical response to treatment with Dimethyl Fumarate.
Figure 7.6	Examples of two multiple sclerosis patients with or without paramagnetic rim lesions.

Figure 7.7	The variant in HIF1A affects the volumetric burden of the paramagnetic rim lesions.
Figure 9.1.1	Principal Component Analysis for the OSR cohort.
Figure 9.3.1	Effect of the adjustment for the seasonal variation on the vitamin D levels.
Figure S2.1	Overview on the work.
Figure S6.1	An extract of the Environmental Questionnaire, section E (Body weight and dietary habits).
Tables	
Table 1.1	Description of the main EDSS steps.
Table 1.2	Currently approved DMT and off-labels drugs in MS.
Table 3.1	Clinical and demographic features of the cohorts included in the genetic study.
Table 3.2	Results of the genome-wide meta-analysis on the NEDA-3 status.
Table 3.3	Gene Ontology (GO) enrichment analysis for the top 100 associated SNPs.
Table 3.4	KEGG pathway analysis for the top 100 associated SNPs.
Table 4.1	Baseline characteristics of the cohort.
Table 4.2.	Top 20 Differentially Methylated Positions (DMP) according to the NEDA/EDA status.
Table 4.3	Differentially Methylated Regions (DMR) in NEDA vs EDA patients.
Table 4.4	Methylation Quantitative-Trait-Loci (mQTL) effect exerted by rs2240656 (C allele) on the 4 CpGs composing the identified DMR in AMH.
Table 4.5	Results of Gene Ontology enrichment analysis.
Table 5.1	Baseline characteristics of the cohort.
Table 5.2	Genetic loci that are associated with serum vitamin D levels variation in controls and with NEDA-3 status.
Table 5.3	Results of the Inverse variance weighted Mendelian Randomization analysis and comparison with different methods.

Table 6.1	Clinical features of the cohort included in the study on environmental factors.
Table 6.2	The impact of age at menarche, oral contraceptive use, menopause, and pregnancy on disease severity.
Table 6.3	Body weight and disease severity assessed by the ARMSS score.
Table 6.4	Alcohol consumption in women and disease severity.
Table 6.5	Alcohol consumption in men and disease severity.
Table 7.1	Clinical and demographic features of the two cohorts involved in genetic association study.
Table 7.2	Top 10 associated SNPs in the OSR cohort and replication in the SWE cohort.
Table 9.1.1	Covariate analysis of clinical and demographic factors towards the NEDA-3 status.
Table 9.2.1	Overview of the QC at the individual- and probe-level on the methylation profile.
Table 9.4.1	Classification of patients into categories for alcohol consumptions.
Table 9.5.1	Inclusion criteria for the patients involved in the genetic association study.
Table S3.1	Results from the GWAS in the OSR cohort.
Table S3.2	Results from the GWAS in the CHUT cohort.
Table S7.1	List of the mapped genes in iron metabolism included in the genetic analysis.
Table S7.2	List of gene sets from Gene Ontology (GO), Human Phenotype (HP) and Kyoto Encyclopedia of Genes and Genome (KEGG) included in the study.

1. Introduction: Multiple sclerosis

1.1 Epidemiology and clinical presentation

Multiple sclerosis (MS) is a chronic immune-mediated disorder of the central nervous system (CNS), characterized by demyelination and neurodegeneration.

Globally, MS is a major cause of disability, especially among young people, with a peak incidence between the age of 20 and 40 years. Its prevalence is approximately 50-300 patients per 100,000 people and varies across different countries, affecting 2-3 million individuals worldwide (Figure 1.1). MS is nearly three times more common in women than in men (Walton et al, 2020).

The clinical manifestations of the disease are diverse, as they depend on the area of the CNS that is affected by the demyelinating lesions, typically considered the hallmark of the disease. Demyelinating lesions can occur in the brain, the spinal cord, and the optic nerves, resulting in a variety of symptoms. The most common are sensory and motor impairment, myelitis, optic neuritis, and bladder dysfunction. At onset, the majority of patients present with a relapsing-remitting (RR) clinical course, which is typically characterized by relapses, monophasic episodes of neurological dysfunction followed by a different degree of recovery, from complete to poor.

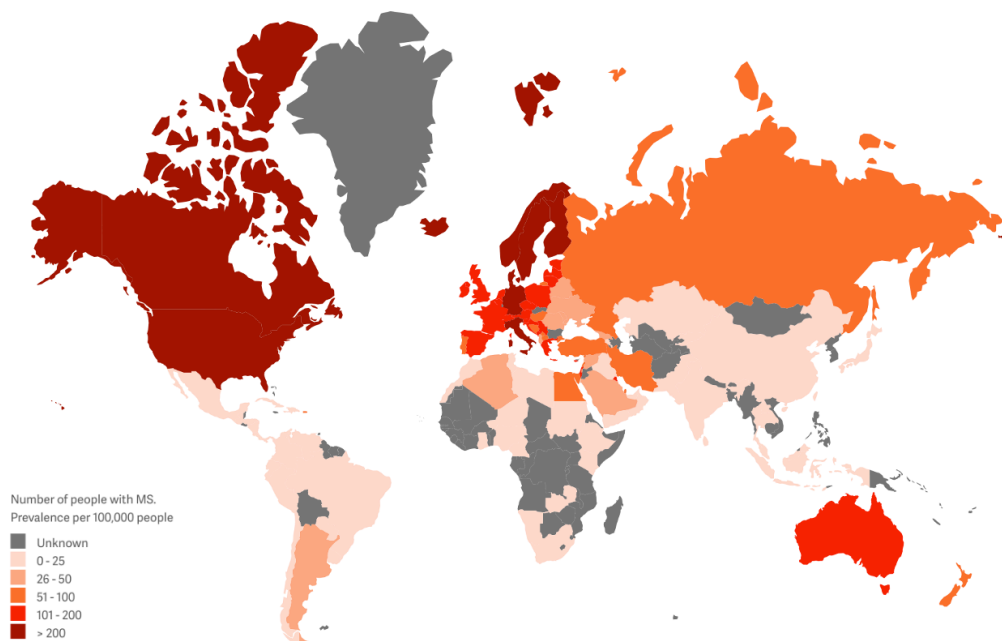


Figure 1.1. Prevalence of MS in the world. From: *Atlas of MS* (<https://www.atlasofms.org/>)

About 10% of MS patients has a primary progressive (PP) course (Koch et al, 2009), which is characterized at onset by a slow and progressive development of disability over time, despite few or no relapses. Patients presenting with a relapsing-remitting course may experience a secondary progression (SP), characterized by a continuous clinical worsening after the initial relapsing phase.

The most widely used tool to quantify disability in MS patients is the Expanded Disability Status Scale (EDSS) score (Kurtzke JF, 1983). It is an ordinal score based on the clinical evaluation of eight functional systems (pyramidal, sensory, visual, bladder and bowel, cerebral, brainstem, cerebellar and other), and it ranges from 0 (no signs of neurological dysfunction) to 10 (death) (Table 1.1). Despite many criticisms, the EDSS score is still the main tool to assess disability progression and severity, both in clinical routine and pharmacological trials (Meyer-Moock et al, 2014).

<i>EDSS</i>	<i>Disability</i>
0.0	Normal neurological exam
1.0	No disability, minimal signs in one FS
2.0	Minimal disability in one FS
3.0	Moderate disability in one FS or mild disability in three or four FS though fully ambulatory.
4.0	Fully ambulatory without aid, self-sufficient, up and about some 12 hours a day despite relatively severe disability; able to walk without aid or rest greater than 500 meters.
5.0	Ambulatory without aid or rest for about 200 meters; disability severe enough to impair full daily activities
6.0	Intermittent or unilateral constant assistance (cane, crutch, brace) required to walk about 100 meters with or without resting
7.0	Unable to walk beyond approximately 5 meters even with aid, essentially restricted to wheelchair; wheels self in standard wheelchair and transfers alone; up and about in wheelchair some 12 hours a day
8.0	Essentially restricted to bed or chair or perambulated in wheelchair, but may be out of bed itself much of the day; retains many self-care functions; generally has effective use of arms
9.0	Helpless bed patient; can communicate and eat
10.0	Death due to MS

Table 1.1. Description of the main EDSS steps.

1.2 Diagnosis

The current diagnostic criteria for MS rely upon the satisfaction of demonstration of both spatial and temporal dissemination of the disease, as a core feature. The last revision of the diagnostic criteria was published in 2018 and allows an accurate diagnosis, based on the combination of clinical, Magnetic Resonance Imaging (MRI) and cerebrospinal fluid (CSF) analysis data (Thompson AJ et al, 2018).

Brain and spinal cord MRI are fundamental for the diagnosis of MS. First, MRI is needed to fulfill the spatial dissemination criterion, which is based on the evidence of demyelinating lesions in at least two out of four areas of the CNS (cortical/juxtacortical, periventricular, infratentorial and spinal cord) (Thompson AJ et al, 2018). Second, the evidence of the simultaneous presence of gadolinium-enhancing and non-enhancing lesions can also demonstrate the dissemination in time, provided that the patient has had at least one clinical attack with symptoms identified by the clinician (Thompson AJ et al, 2018). The dissemination in time can also be demonstrated by the presence of oligoclonal bands in the CSF or the history of at least 2 relapses (Thompson AJ et al, 2018). MRI is also crucial in identifying other conditions that may mimic MS and that must be carefully ruled out before making diagnosis of MS (Filippi et al, 2019).

Research is now focusing on improving the performance of the diagnostic criteria for MS, for example including symptomatic optic nerve involvement or the presence of lesions showing a central vein sign in the determination of dissemination in space (Brownlee et al, 2018; Sinnecker et al, 2019).

1.3 Pathogenesis

MS has a complex multi-factorial pathogenesis, which is still largely unknown, despite many recent advancements. Genetic and environmental factors play a central role in determining both the susceptibility to the disease and its manifestations, which involve inflammatory and neurodegenerative mechanisms (Filippi et al, 2018).

Briefly, the migration of autoreactive immune cells from the periphery to the CNS and the initiation of an inflammatory process against the molecular components of myelin has been considered for a long time the core feature of the pathobiology of the disease (Bar-Or et al, 2021) (Figure 1.2). In a more modern view, such mechanisms integrate with an underlying neurodegenerative process, which seems to arise early in MS and to proceed

in parallel and in combination with the inflammatory counterpart (Bar-Or et al, 2021). Crucial is the role of microglial cells that react to inflammation, become overly activated contribute to maintain a pro-inflammatory and neurodegenerative status (Absinta et al, 2021).

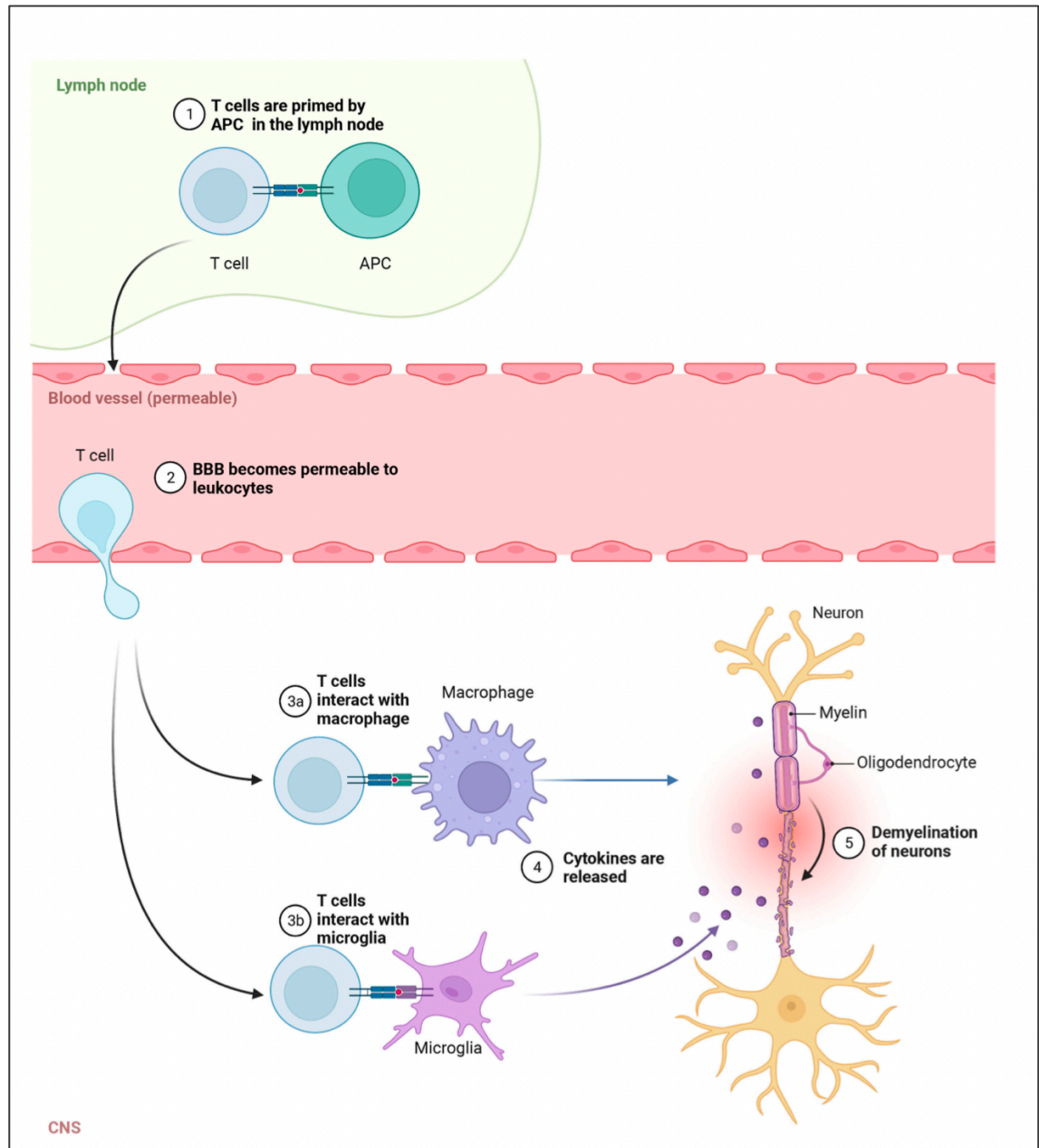


Figure 1.2. Mechanisms of myelin damage in multiple sclerosis. From Liu et al, 2022.

1.3.1 Genetic factors

In the past years, different studies have given an important contribution in assessing the role of genetic factors in the susceptibility to MS (Goris et al, 2022). On this purpose, the latest and largest study was published in 2019 by the International Multiple Sclerosis Genetics Consortium (IMSGC) (International Multiple Sclerosis Genetics Consortium, 2019). The study investigated the effect of millions of Single Nucleotide Polymorphisms (SNPs) on the susceptibility to develop the disease in 47,351 patients with MS and 68,248 controls of European ancestry. Thanks to this effort, 233 independent genetic loci have been found to impact the risk of developing MS, including 32 risk variants in the Human-Leukocyte Antigen (HLA) system and 1 risk variant in chromosome X. This study included SNPs which were defined as common in the general population, having a Minor Allele Frequency (MAF) $\geq 5\%$. Cumulatively, such variants explain about 40% the heritability of MS. Variants in the HLA system are the strongest genetic factor that has merged so far, since the first studies that have been published (Goris et al, 2022). HLA class I and II alleles code for molecules that are involved in the presentation of antigens to CD8 and CD4 lymphocytes, the activation of which seems to be an important step in the cascade of MS pathogenesis. In particular, HLA-DRB1*1501, tagged by the SNP rs3135388 (De Bakker, McVean et al., 2006), seems to exert the biggest risk, with an Odds Ratio (OR) of approximately 3 (Brynedal et al, 2007; Sawcer et al, 2011). Conversely, the HLA-A*02 allele has shown a protective effect (Sawcer et al, 2011). In a previous study from the IMSGC the role of variants with a low frequency (1-5%) were investigated, and significant associations emerged (International Multiple Sclerosis Genetics Consortium, 2018), even though the relative contribution of these variants seems more modest, also considering that larger cohorts are needed to successfully identify more loci. Many reports on small cohorts tried to determine the role of rare SNPs (MAF < 1%), but in lack of proper validation studies the contribution of these variants still remains unclear (Harding et al, 2019).

1.3.2 Environmental factors

The impact of some environmental factors on the risk of developing MS has been explored by many studies. Overall, the environment is fundamental in the pathobiology

of the disease, considering that only part of the risk of MS has been attributed to genetic predisposition (Olsson et al, 2017).

1.3.2.1 Epstein-Barr Virus

Among the environmental factors, particular attention has been gathered by the Epstein-Barr Virus (EBV) infection. A consistent body evidence has shown an association between EBV seropositivity and MS, as nearly all the patients have detectable IgG antibodies against EBV (Houen et al, 2020). A recent study has assessed that the risk of developing MS in 955 people diagnosed with MS after EBV infection was increased by 32-fold after EBV infection, while the risk was not affected when infection from other viruses occurred (Bjornevik et al, 2022). In the same work, the levels of serum neurofilament light chain (NFL), a recognized marker of ongoing inflammation and neuroaxonal degeneration, started to increase only after EBV infection. These data provide strong evidence that EBV is a very important factor in determining MS, even though the biological mechanisms underlying this association are still not known. EBV resides in a latent state in B cells, and it has been suggested that this factor could contribute to activate B lymphocytes through mechanisms of molecular mimicry (Houen et al, 2020). Conversely, studies showing invasion of the CNS of EBV-infected cells and the presence of EBV positive cells in MS lesions have yielded controversial results (Sargsyan et al, 2010; Hassani et al, 2018). Therefore, despite the epidemiologic correlation, the causal mechanism underlying EBV infection and the increased risk of MS is still unclear.

1.3.2.2 Cigarette smoking

Smoking, both active and passive, is an important factor for MS, in a dose-dependent manner (Ghadirian et al, 2001; Hedström et al, 2011). The main hypothesis is that peripheral activation of lymphocytes occurs in lungs in patients who smoke as a consequence of chronic pulmonary irritation. Autoreactive lymphocytes then migrate into the CNS and initiate the inflammatory process that leads to demyelination and formation of MS lesions. This mechanism is also supported by the finding that the use of oral tobacco, very popular in Scandinavian countries, was not associated with increased risk of MS (Hedström et al, 1999). Conversely, its content in nicotine, which showed

neuroprotective effect in preclinical models, may reduce the risk of MS, but larger studies are needed to validate this result (Nizri et al, 2009).

Interestingly, smoking dramatically increases the risk of MS in carriers of the HLA risk variants in a synergistic way. People who smoke, who are carriers of the risk allele HLA-DRB1*15:01 and who lack the protective HLA-A*02 allele have an OR of developing MS of ~14 (Hedström et al, 2011).

1.3.2.3 Vitamin D

Higher vitamin D levels, favored by greater sun exposure, reduce the risk of MS (Munger et al, 2006; Salzer et al, 2012), partly explaining why MS is more common in Northern European countries (together with a higher prevalence of HLA-DRB1 risk allele in these populations). Higher vitamin D levels were also found to be associated to a better outcome in terms of disease activity and disability progression in patients treated with interferon beta-1b (Fitzgerald et al, 2015).

Through a major regulatory role on gene expression, vitamin D exerts immunomodulatory effects in both adaptive and innate immunity (Gombash et al, 2022). Some studies have suggested a neuroprotective effect of vitamin D, given its role in the regulation of calcium intake in neurons (Hausler et al, 2019) and in the reduction of oxidative stress through the Nrf2 transcription factor (Nachliely et al, 2019).

1.3.2.4 Gut microbiota, obesity, and other factors

The term ‘gut microbiota’ is referred to commensal and pathogenic bacteria that colonize the intestine. Quite recently, an increasing body of evidence is promoting its role in MS. In the preclinical setting, mice raised in a germ-free environment were resistant to the induction of the Experimental Autoimmune Encephalomyelitis (EAE), the most widely used animal model for MS. EAE was then triggered when microorganisms were later introduced with feeding (Berer et al, 2011). Subsequent studies were successful in showing that the gut microbiota is significantly different in MS compared to controls (Lee et al, 2010; Miraza et al, 2017; Berer et al, 2017), and in RR-MS versus progressive MS (Cox et al, 2021).

The driving hypothesis is that the microbiota is able to regulate the activity and the maturation of both T and B lymphocytes, in addition to a stimulatory effect on the

production of serotonin, which in turns regulates the immune system itself (Cosorich et al, 2017; Correale et al, 2022). The gut microbiota is influenced by genetic and dietary factors, and therefore an effort in assessing the role of specific dietary interventions as potential therapeutic options for MS is ongoing (Correale et al, 2022).

Obesity also seems to have a role in MS. Multiple studies have pointed out that obesity early during adolescence is associated with a greater risk of developing MS (Munger et al, 2013; Langer-Gould et al, 203), also supported by Mendelian randomization studies (Mokry et al, 2016; Gianfrancesco et al, 2017). Biologically, obesity can act on multiple pathways that either sustain a pro-inflammatory environment or lead to decreased vitamin D levels, which than alters immunomodulatory functions.

Data showing the effect of other environmental factors, like alcohol or caffeine consumption, are to date more controversial, as in small-medium studies opposite effects have been found (Olsson et al, 2017).

1.4 Treatment

Over the years, patients and clinicians have seen a great expansion in the number of pharmacological options to treat MS. The use of Disease-Modifying Treatments (DMT), which have been proven effective in dramatically reducing the occurrence of new relapses, has led to an improvement in patients' quality of life (Jongen, 2017). The therapeutic approach is modelled on various individual characteristics, considering clinical and radiological disease activity, prognostic factors, comorbidities, and patient's compliance.

Different lines of treatment are available for MS, with increasing level of efficacy but, alongside, side effects and risks. Therefore, two strategies to treat MS at onset have been largely debated: the escalating and the induction approach (Prosperini et al, 2020). In the escalating approach, patients may be exposed to the risk of assuming drugs with under-target efficacy, despite less frequently developing complications and side effects from the treatment. In the induction approach, the clinicians identify patients at a higher risk of short and long-term disability and start early with a high-efficacy treatment, controlling for the potential risk. The rationale for this approach is that disease activity in early relapsing phases predicts long-term disability and represents a strong negative prognostic factor (Scalfari et al, 2014; Fisniku et al, 2008).

DRUG	MECHANISM OF ACTION	LINE	TARGET
Interferons	Regulation of adaptive immunity, blood-brain barrier permeability	I	RR-MS
GA	Anti-inflammatory shift from Th1 to Th2 lymphocytes	I	RR-MS
Teriflunomide	Inhibition of dihydroorotate dehydrogenase and reduced immune cell growth	I	RR-MS
DMF	Inhibition of Nrf-2 and NF-kB pathways, reduction of oxidative and inflammatory stress	I	RR-MS
Natalizumab	Blocking of alpha-4 integrin; impaired lymphocytes and monocytes migration into the CNS	II	RR-MS
S1P modulators	Reduction of lymphocytes migration from the lymph nodes to the CNS	II	RR-MS ¹ SP-MS ²
Cladribine	Purine analogue, depletion of lymphocytes	II	RR-MS
Anti-CD20	Depletion of CD20+ B lymphocytes	II	RR-MS PP-MS ³
Mitoxantrone	Type II topoisomerase inhibitor	II	RR-MS
Off-Label:		-	-
<i>Azathioprine</i>	Inhibition of purine synthesis		
<i>Rituximab</i>	Depletion of CD20+ B lymphocytes		
<i>Cyclophosphamide</i>	Alkylating agent		

Table 1.2. IFN=Interferon. Drugs in the Interferons group are: IFN-beta-1a (Avonex, Rebif-22, Rebif-44, Plegridy), IFN-beta-1b (Betaferon, Extavia). GA=Glatiramer acetate (Copaxone). Drugs in the S1P group are: fingolimod (Gylenia), Siponimod (Mayzent), Ozanimod (Zeposia). 1. Fingolimod and Ozanimod are approved for RR-MS. 2. Siponimod is approved for SP-MS. Drugs of the anti-CD20 group are: Ocrelizumab (Ocrevus) and Ofatumumab (Kesimpta). 3. Ocrelizumab is approved for PP-MS.

Even though recent evidence has suggested a better long-term outcome in patients treated with early intensive therapy versus first-line moderate-efficacy DMTs (Harding et al, 2019), the debate on the topic is still open.

A schematic overview on the currently approved DMT and the off-label drugs used in MS, their mechanism of action and their target is reported in Table 1.2.

1.5 Disease activity in MS: key elements

As mentioned, MS has a very heterogenous clinical course in terms of disease activity. Some patients present with a very low burden of relapses and lesions over the years and accumulate no or mild disability, while in others the disease has a much more pronounced activity that leads to a rapid deterioration of patients' quality of life (Díaz et al, 2019). Hence, markers that can reliably give indications on the probability of future accumulation of disability and that can be used early, after the diagnosis of MS is formulated, are strongly needed. Finding molecular markers may also give an important insight on the biological mechanisms underlying the manifestations of the disease, which can prompt new therapeutic options.

Demographic and clinical prognostic factors (as age at onset, gender, disability level at onset, etc.) have been in use for a long time, and their role in the clinical practice is still valid (Tintore et al, 2015). However, they are not able to detect all the aspects involved in the disease. The ideal markers could aid the clinicians when deciding on treatment, with the goal of minimizing the risk of under or overestimation of the potential consequences of the disease or of the treatments.

The first step in order to accomplish this goal is to have reliable and powerful outcomes that capture the manifestations of the disease, and its related disability. The availability of effective and reliable outcomes of disease activity is key to achieve meaningful results in a clinical or translational study.

Over the past few years, the definition of disease activity has evolved, along with new acquisitions coming from basic, clinical, and translational research. In a more classical view, the term 'disease activity' referred to the more easily measurable manifestations of the disease, as the clinically defined relapses. The first outcomes used in clinical trials were often unidimensional, as for example the annualized relapse rate (ARR) (Nicholas et al, 2012). Although the ARR gives precious information, this kind of outcome is limited

in terms of sensitivity. With the advent of MRI and its routine use in the diagnosis and follow-up of MS, more sensitive measures of disease activity were introduced in clinical trials, targeting, for example, the number of new or gadolinium-enhancing lesions. Most importantly, composite outcomes of disease activity were then introduced, integrating different levels of information from the clinical evaluation and the MRI examinations. In this regard, the No Evidence of Disease Activity (NEDA) status has emerged as a widely accepted measure of disease activity, and it has been implied in many clinical trials (Banwell et al, 2013). In its more modern version, the NEDA status is often translated into the NEDA-3 status, as its definition is composed by three dimensions: relapses, EDSS progression and MRI activity (Banwell et al, 2013). Many researchers have also recently suggested to expand the NEDA-3 status into a NEDA-4, adding measures of brain atrophy at MRI (Kappos et al, 2016; Kappos et al, 2021).

Indeed, nowadays, the assumption that disease progression, clearly evident from the clinical point of view, is silent at the imaging level as well, has radically changed. Despite the absence of new lesions at conventional MRI examinations, consistent brain and spinal cord changes are actually detected, if investigated with high sensitivity methods, and provide a robust measure of underlying degeneration (Rocca et al, 2017). Moreover, a consistent body of evidence has shown that, in some patients, neurodegeneration and brain volume loss start early, even in the RR course (Eshaghi et al, 2018; Cagol et al, 2022).

After establishing that measuring brain atrophy was fundamental in MS, the scenario radically evolved again over the past few years, thanks to the identification of the paramagnetic rim lesions (PRL) by the means of advanced susceptibility imaging techniques (Absinta et al, 2016; Absinta et al, 2018) (Figure 1.3). The PRL represent the in vivo correlate of the chronic active lesions, that are known from pathological studies to be more frequent in patients with progressive MS (Kuhlmann et al, 2017). The burden of PRL is associated with disability measures in patients with MS (Absinta et al, 2019), mirroring a process characterized by silent inflammation which is not detected by conventional MRI. The silent inflammation is nevertheless a manifestation of the disease and can be seen as part of the definition of disease activity, even though it is not characterized by discrete relapses that are easily recognized.

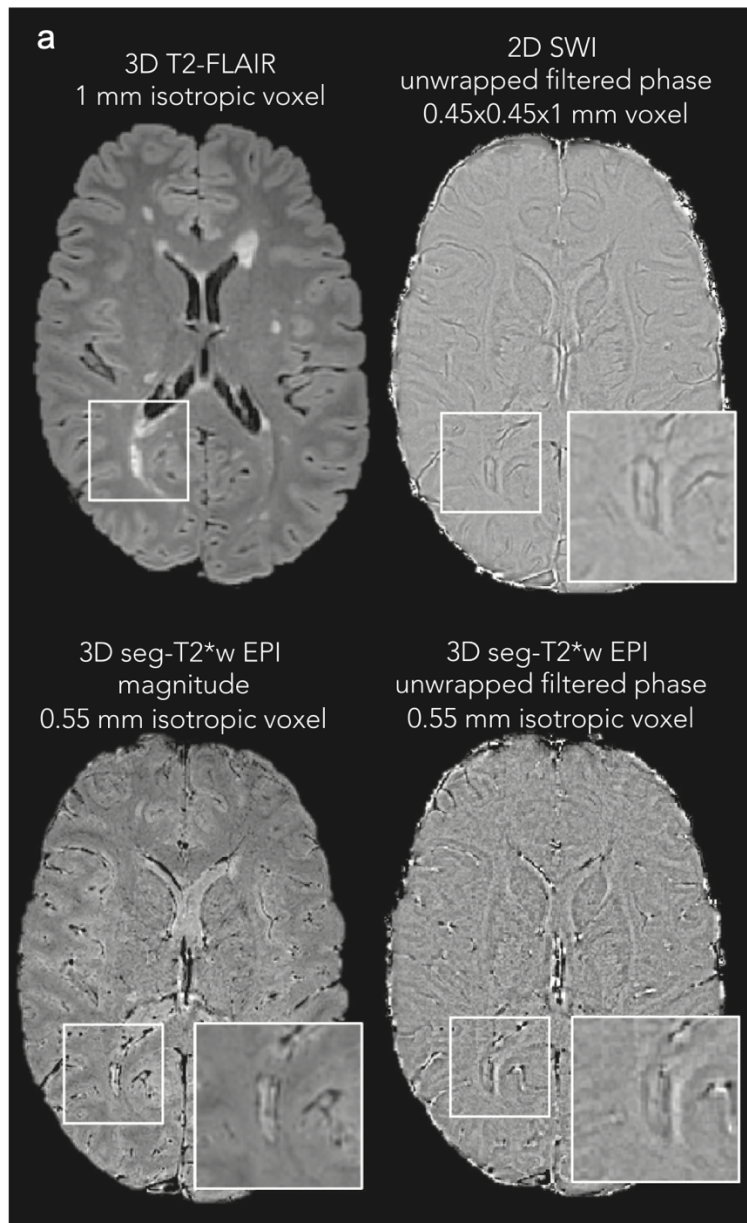


Figure 1.3. Paramagnetic Rim Lesions, adapted from Martire et al, 2022. Example of a paramagnetic rim lesion (PRL) at 3-Tesla MRI in a patient with MS. Information on the different MRI sequences and image resolution is reported by the authors. FLAIR = Fluid Attenuated Inversion Recovery; SWI = Susceptibility Weighted Imaging; EPI = Echo Planar Imaging.

Nowadays, many drugs that hamper inflammation in the RR course of MS are available, and they are effective in limiting the number of relapses or measures of disease activity at conventional MRI. Conversely, very little is known whether already available treatments have an effect also on the silent inflammation which is captured by the PRL. Therefore, it has been suggested that the PRL will rapidly become a fundamental outcome of disease activity in future trials, and some effort in this regard is already ongoing

(Martire et al, 2022). Interesting measures will be, as an example, the proportion of PRL in which the paramagnetic rim fades over time after the beginning of a pharmacological treatment, or the mean volume of the PRL, as some of these lesions tend to expand over the time due to the underlying inflammatory process and are referred to a slowly expanding lesions (Elliott et al, 2019).

Another important concept, that is strictly related with disease activity and disability progression, is disease severity. Disease severity ranks the disability that affects patients and therefore is a comprehensive measure that reflects, at a more functionally oriented level, the damage due to underlying disease activity (Kister et al, 2020). As mentioned, the most widely accepted tool to rate disability in MS is the EDSS score. This score faces many pitfalls, mainly due to non-linearity of its measurement (Kister et al, 2020). To overcome this limit, the use of other severity scores with a linear distribution has been suggested. Among the most important severity scores in MS, it is worth mentioning the Multiple Sclerosis Severity Score (MSSS), which ranks patients' disability based on EDSS score adjusted for disease duration (Roxburgh et al, 2005), and the Age-Related Multiple Sclerosis Severity (ARMSS) score, which relies on EDSS and age (Manouchehrinia et al, 2017). The use of ARMSS instead of MSSS allows to prevent a potential bias due to the uncertainty on when exactly the disease started in many patients, also considering that most likely the biological cascade of MS is activated earlier than evident clinical manifestations occur (Manouchehrinia et al, 2017).

In conclusion, our view is that the clinical diversity of the disease is sustained by different shades of the same integrated biological process. It is clear that one single comprehensive outcome measure does not exist; therefore, in the present work, we will imply different kinds of outcomes to study the same entity from different angles.

2. Aims of the work

Understanding the biological basis of a disease is crucial for a better management of patients, the successful development of new therapeutic options (Dugger et al, 2018) or for the repurposing of already available drugs (Pushpakom et al, 2019).

As we discussed above, MS is heterogenous in terms of clinical manifestations, and very little is known about the biological mechanisms that underlie this heterogeneity. Gaining such knowledge would be extremely important for different reasons: 1) to have meaningful markers that could help a tailored treatment; 2) to drive future drug development, especially for progressive MS, for which the treatment is still largely an unmet need.

Therefore, the overall aim of this work is to provide an insight on the mechanisms that contribute to the clinical heterogeneity that characterizes MS, focusing on measures of disease activity, severity, and progression. Our approach will include different kinds of molecular and environmental information, and it will be modelled on the basis of an underlying clinical reasoning. An overview of the design of the present work is shown in Figure S2.1.

Specifically, our aims are:

1. To assess the molecular determinants (genetic, epigenetic and vitamin D levels) of disease activity assessed by the NEDA-3 status;
2. To investigate whether the exposure to specific environmental factors early during lifetime affects future disease severity;
3. To study potential genetic factors driving chronic silent inflammation in progressive MS.

RESULTS

3. Study 1: Genetic factors underlying disease activity

In this first study, we aim to explore the genetic factors that could potentially influence the risk of disease activity in MS. Despite many works that have assessed the genetic determinants of the susceptibility to MS (IMSGC, 2011; IMSGC, 2019), very few papers investigated measures of disease activity, and they were mainly focused on relapse hazard only (Hilven et al, 2018; Vandebergh et al, 2021).

Herein, we adopt the No Evidence of Disease Activity (NEDA-3) status, a well-established composite outcome to identify disease activity in MS (Banwell et al, 2013). Specifically, we will imply the NEDA-3 status to assess whether there are genetic factors that affect the risk of having signs of disease activity at 2-years follow-up from the beginning of a new first-line DMT, meta-analyzing an Italian cohort of 1,183 patients and a French cohort of 299 patients. The clinical features of the two cohorts are reported in Table 3.1. Inclusion of the patients is discussed in the Methods section (Chapter 9.1.1)

	<i>OSR</i>	<i>CHUT</i>	<i>p-value</i>
<i>No.</i>	1,183	299	-
<i>F/M ratio</i>	2.25 (819/364)	3.27 (229/70)	<0.001
<i>Age (mean, SD)</i>	34.56 (9.67)	36.20 (10.32)	0.014
<i>Disease duration</i>	5.42 (6.06)	5.90 (6.56)	0.84
<i>Median EDSS at BL (IQR)</i>	1.5 (1.0-2.0)	1.5 (1.0-2.5)	0.34
<i>Relapses before BL (mean, SD)</i>	1.38 (0.98)	1.45 (1.13)	0.71
<i>DMT started at BL</i>	IFN 774 (65%) GA 327 (28%) DMF 69 (6%) TERI 12 (1%)	IFN 156 (53%) GA 46 (15%) DMF 54 (18%) TERI 39 (14%)	-
<i>NEDA-3 at 2 years</i>	EDA 782 (66%) NEDA 401 (34%)	EDA 223 (75%) NEDA 76 (25%)	0.0062

Table 3.1. Clinical and demographic features of the cohorts included in the genetic study. Differences between the two cohorts were tested by the Mann-Whitney or chi-square test, as appropriate, and the p-value is reported in the rightmost column. F/M: Female/Male. SD=standard deviation. EDSS = Expanded Disability Status Scale score. IQR = interquartile range. BL = baseline. DMT = Disease-Modifying Treatment. IFN = interferon. GA = Glatiramer acetate. DMF = Dimethyl Fumarate. TERI = Teriflunomide. EDA = Evidence of Disease Activity.

3.1 Results

3.1.1 Meta-analysis results

After Quality Controls (QC), first we performed the two preliminary single-cohort GWAS for OSR and for CHUT, finding no genome-wide significant association. Different Single Nucleotide Polymorphisms (SNPs) showed a suggestive association with the NEDA-3 status (Appendix: Supplementary Results, Table S3.1 and Table S3.2).

We then proceeded with a fixed-effect meta-analysis to increase statistical power, taking advantage of the combination of the two cohorts. We meta-analyzed 3,948,158 SNPs that were common to the two cohorts, in a total of 1,408 patients. A quantile-quantile (QQ) plot showing the genomic inflation factor is shown in Figure 3.1, suggesting no significant population stratification ($\lambda = 1.012$).

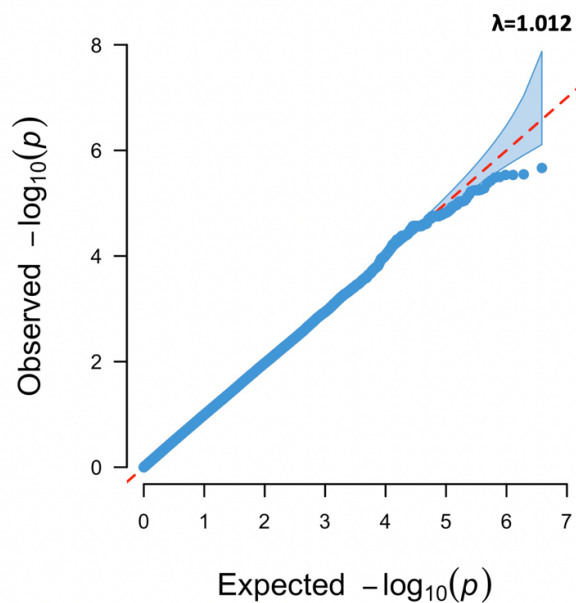


Figure 3.1. QQ-plot showing the genomic inflation factor (λ) for the meta-analysis.

In the meta-analysis, we did not find any genome-wide significant association using a Bonferroni correction to account for multiple testing ($p < 5e-08$), but many variants showed a p-value which was suggestive for association with the NEDA-3 status ($p < 1e-05$). The results of the meta-analysis are reported in Table 3.2 and a Manhattan plot is shown in Figure 3.2. The heterogeneity level explored by Cochran's Q and I^2 was overall acceptable, considered the difference in sample size between the two cohorts (Table 3.2).

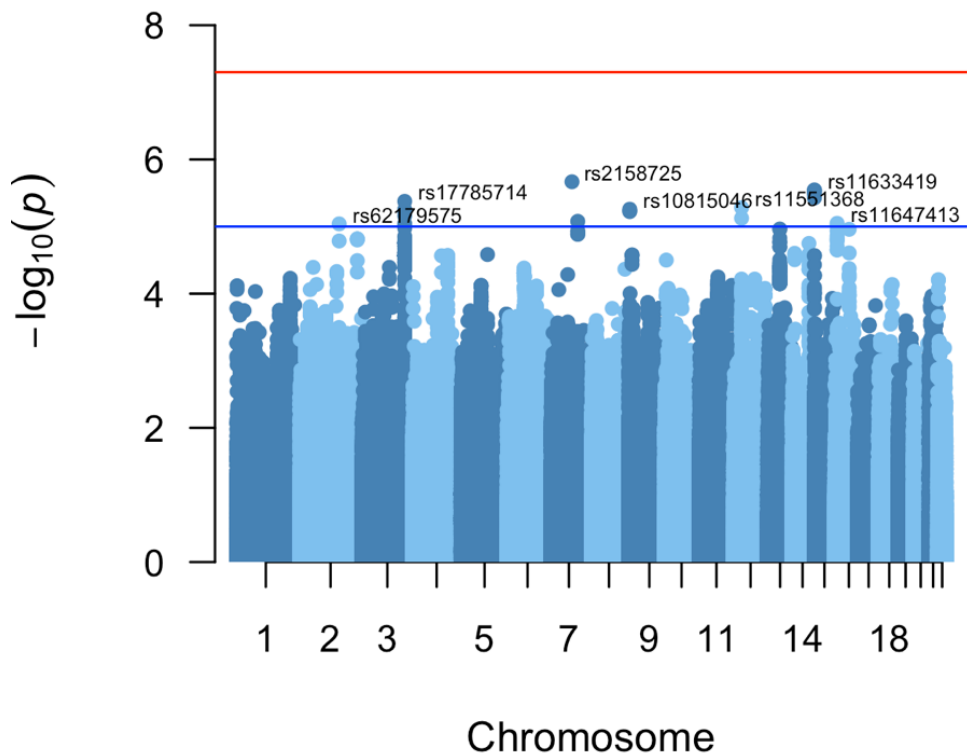


Figure 3.2. Manhattan plot of the genome-wide meta-analysis between OSR and CHUT. Red line = genome-wide significant threshold (p -value $< 5e-08$); blue line = threshold for a suggestive association (p -value $< 1e-05$)

We then used the OpenTarget *VariantToGene* pipeline to assign genes to variants, in a way which is both position- and function-informed. The top-associated gene for each of the top-associated variant is reported in the rightmost column in Table 3.2.

All the technical details regarding genotype imputation, single-cohort QC, genome-wide association studies, meta-analysis and the OpenTarget pipeline are described in the Methods section (Chapter 9.1).

The top-associated signal, tagged by rs2158725 ($p=2.15e-06$; $OR_{EDA}=0.57$), maps to the Semaphorin-3E (*SEMA3E*) gene. The Minor Allele Frequency (MAF) of rs2158725 is 0.14 in the European population (Karczewski et al, 2020), and the variant is known to exert a functional effect on the *SEMA3E* gene, through the modulation of DNA methylation (GoDMC, 2021). The role of the semaphorins, fundamental regulators of the immune system and myelin repair, and the functional impact of the variant are discussed more extensively in Chapter 3.2.

<i>CHR</i>	<i>BP</i>	<i>SNP</i>	<i>AI</i>	<i>P</i>	<i>OR_{EDA}</i>	<i>Q</i>	<i>I</i>	<i>Gene</i>
7	83297241	rs2158725	C	2.15E-06	0.57	0.24	27.96	SEMA3E
15	23779479	rs11633419	T	2.84E-06	0.68	0.11	61.4	MKRN3
3	167897639	rs17785714	T	4.20E-06	0.45	0.24	27.16	GOLIM4
12	31825120	rs11551368	C	5.23E-06	1.80	0.77	0	RESF1
9	4641755	rs10815046	T	5.51E-06	0.59	0.46	0	PLPP6
7	105954740	rs2704966	T	8.35E-06	0.67	0.31	3.26	NAMPT
16	10424872	rs11647413	G	8.95E-06	1.50	0.66	0	ATF7IP2
2	153953265	rs62179575	T	9.13E-06	0.65	0.25	24.44	PRPF40A

Table 3.2. Results of the genome-wide meta-analysis on the NEDA-3 status. The summary statistics for lead variant of the top associated genetic loci ($p < 1e-05$) are reported, together with the mapped gene using the OpenTarget pipeline (see text for further detail). BP = base-pair positions (reference GRCh37/hg19). AI = effect allele; P = p-value; OR_{EDA} = Odds Ratio for EDA status. Q = Cochran's Q; I = I^2 statistics for heterogeneity; GENE = UCSC RefSeq gene symbol of the top-associated gene at the OpenTarget VariantsToGene analysis.

The second top-associated signal (lead SNP: rs11633419; $p=2.84e-06$; $OR_{EDA}=0.68$) involved the *MKRN3* (Makorin Ring Finger Protein 3) gene, which has been shown to regulate the onset of puberty in mammals, as it is expressed in brain and acts on the Gonadotropin-Releasing hormone during childhood (Abreu et al 2020; Li et al, 2020).

Among the variants showing a suggestive association, also the signal in chromosome 7 mapping to the *NAMPT* (Nicotinamide Phosphoribosyltransferase) gene may be very promising, since *NAMPT* inhibition was suggested as a target to enhance neurological recovery in MS (Li et al, 2016). The lead variant (rs2704966; $p=8.35e-06$; $OR_{EDA}=0.67$) is located in an enhancer region of *NAMPT* and has a reported frequency of 0.33 in the European population (Karczewski et al, 2020). The variant is known to exert an expression-QTL (eQTL) and mQTL effect on blood, as discussed in Chapter 3.2.

3.1.2 Gene Ontology enrichment and pathway analysis

To have an overview on the potential biological mechanisms underlying different disease activity levels, we run an enrichment and pathway analysis, as described in the Methods. We found a significant enrichment ratio for the GO:0034341 term ('*Response to interferon gamma*'), despite adjustment for multiple testing. Notably, some other interesting terms, that did not pass the False Discovery Rate (FDR) < 0.05 threshold for

multiple testing, were also nominally enriched, as ‘*Cell chemotaxis*’, ‘*Homotypic cell-cell adhesion*’, ‘*Integrin-mediated signaling pathway*’ and ‘*Neuroinflammatory response*’ (Table 3.3 and Figure 3.3).

Interestingly, most of the same genes that belong to the ‘*Response to interferon gamma*’ term (*JAK2*, *CX3CL1*, *CCL22*, *CCL17*), were also part of the ‘*Chemokine signaling*’ pathway, that resulted significant when the same top 100 genes were used as input in an analysis involving pathways from the Kyoto Encyclopedia of Genes and Genome (KEGG) (FDR-adjusted p-value=0.0187; enrichment ratio = 8.47). The results of the pathway analysis are reported in Table 3.4.

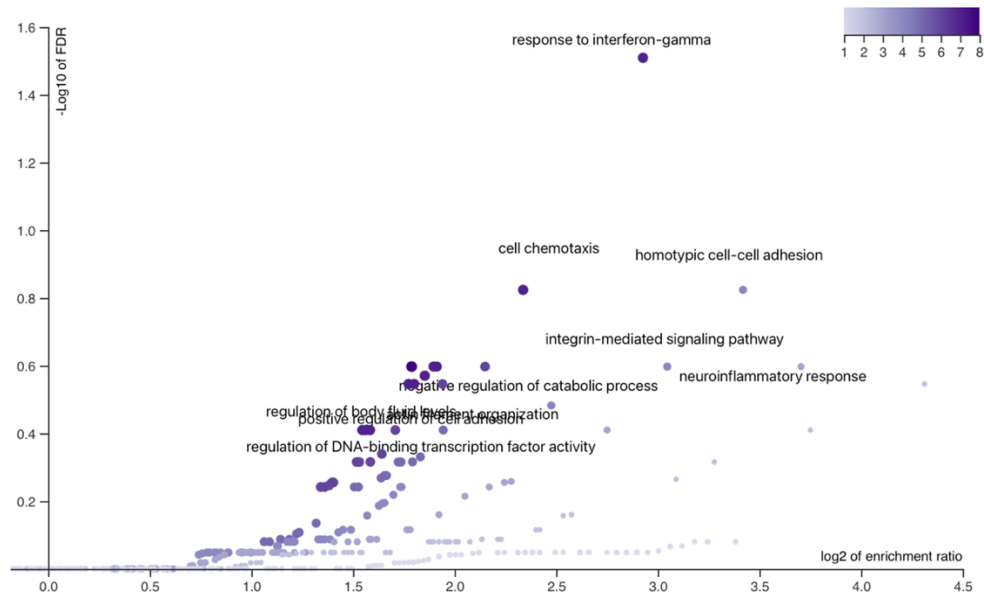


Figure 3.3. Volcano plot showing the results of the Gene Ontology enrichment analysis. Each dot represents a Gene Ontology (GO) term, and the color scale reflects the significance of the p-value after FDR adjustment (reported on a $-\log_{10}$ scale on the Y axis). On the X axis the \log_2 of the enrichment ratio for the GO term is reported.

<i>GO term</i>	<i>description</i>	<i>OL</i>	<i>exp</i>	<i>ER</i>	<i>p-val</i>	<i>p.fdr</i>	<i>genes</i>
<i>GO:0034341</i>	response to interferon-gamma	7/192	0.92	7.60	3.64E-05	0.031	JAK2;NLRC5;CX3CL1;CCL22; CCL17;CD47;CAMK2D
<i>GO:0060326</i>	cell chemotaxis	7/289	1.39	5.05	4.59E-04	0.150	PIK3CG;CX3CL1;CCL22;CCL17; PRKD1;PRKCQ;SBDS
<i>GO:0034109</i>	homotypic cell-cell adhesion	4/78	0.37	10.70	5.30E-04	0.150	PIK3CG;SERPINE2;SLC7A11; PRKCQ
<i>GO:0007229</i>	integrin-mediated signaling pathway	4/101	0.48	8.26	1.40E-03	0.253	EMP2;CUL3;PRKD1;CD47
<i>GO:0150076</i>	neuroinflammatory response	3/48	0.23	13.04	1.56E-03	0.253	NAMPT;CX3CL1;ADCY8
<i>GO:0050878</i>	regulation of body fluid levels	8/483	2.32	3.45	2.12E-03	0.253	JAK2;AK3;PIK3CG;EMP2;ADCY8 SERPINE2;SLC7A11;PRKCQ;
<i>GO:0009895</i>	negative regulation of catabolic process	6/282	1.35	4.44	2.30E-03	0.253	ETFBKMT;PIK3CG;NAMPT; GRIN2A;SERPINE2;SCFD1
<i>GO:0007015</i>	actin filament organization	7/388	1.86	3.76	2.53E-03	0.253	MAGEL2;PDCD10;JAK2;EMP2; CX3CL1;CUL3;CD47
<i>GO:0045785</i>	positive regulation of cell adhesion	7/392	1.88	3.72	2.68E-03	0.253	JAK2;EMP2;ADGRG1;CX3CL1; PRKCQ;HHLA2;CD47

Table 3.3. Gene Ontology (GO) enrichment analysis for the top 100 associated SNPs. *OL* = number of genes overlapping/size of the GO term; *exp* = expected overlap; *ER* = Enrichment Ratio; *p-val* = nominal *p*-value; *p.fdr* = FDR-adjusted *p*-value; *genes* = genes overlapping.

<i>geneSet</i>	<i>description</i>	<i>OL</i>	<i>Exp</i>	<i>ER</i>	<i>p-val</i>	<i>p.fdr</i>	<i>genes</i>
<i>hsa04062</i>	Chemokine signaling pathway	6/189	0.71	8.47	5.74E-05	0.019	JAK2;PIK3CG;CX3CL1;CCL22;CCL17;ADCY8
<i>hsa04725</i>	Cholinergic synapse	4/112	0.42	9.53	7.43E-04	0.121	JAK2;PIK3CG;ADCY8;CAMK2D
<i>hsa04720</i>	Long-term potentiation	3/67	0.25	11.94	1.93E-03	0.209	GRIN2A;ADCY8;CAMK2D
<i>hsa04713</i>	Circadian entrainment	3/96	0.36	8.34	5.34E-03	0.282	GRIN2A;ADCY8;CAMK2D
<i>hsa04925</i>	Aldosterone synthesis and secretion	3/96	0.36	8.34	5.34E-03	0.282	PRKD1;ADCY8;CAMK2D
<i>hsa04750</i>	Inflammatory mediator regulation of TRP channels	3/99	0.37	8.08	5.82E-03	0.282	PRKCQ;ADCY8;CAMK2D
<i>hsa04024</i>	cAMP signaling pathway	4/199	0.75	5.36	6.06E-03	0.282	GRIN2A;CNGB1;ADCY8;CAMK2D
<i>hsa04724</i>	Glutamatergic synapse	3/114	0.43	7.02	8.59E-03	0.350	SLC1A1;GRIN2A;ADCY8
<i>hsa04261</i>	Adrenergic signaling in cardiomyocytes	3/144	0.54	5.56	1.62E-02	0.585	PIK3CG;ADCY8;CAMK2D
<i>hsa04921</i>	Oxytocin signaling pathway	3/152	0.57	5.26	1.87E-02	0.608	PIK3CG;ADCY8;CAMK2D

Table 3.4. KEGG pathway analysis for the top 100 associated SNPs. *OL* = number of genes overlapping/size of the pathway; *exp* = expected overlap; *ER* = Enrichment Ratio; *p-val* = nominal *p*-value; *p.fdr* = FDR-adjusted *p*-value; *genes* = genes overlapping

3.2 Considerations

The genetic drivers of the heterogeneity that characterizes disease activity in MS have been little explored by previous studies. The main reason lies behind the significant sample size that is needed to yield confident results when testing genome-wide associations. International consortia have reached extraordinary advancements in studying the genetic risk of developing MS, taking advantage of meta-analysis of several cohorts from different centers across the world (IMSGC, 2019). Performing the same kind of studies in the context of a multicentric consortium is much more problematic. Exploring phenotypes that go beyond the susceptibility to the disease needs a much more consistent effort, with the harmonization between clinical data across centers representing the biggest pitfall. In this context, the present work, taking advantage of two independent European cohorts, tries to explore possible genetic determinants of disease activity in MS, using the NEDA-3 status as a measure, a composite outcome used in many clinical trials (Banwell et al, 2013).

The top associated signal in the meta-analysis involves the *SEMA3E* gene, with the minor allele C for the lead variant rs2158725 exerting a protective effect towards disease activity. Literature data support a functional impact of the lead variant. In a multicentric effort involving about 36,000 controls (GoDMC, 2021), rs2158725 exerted a mQTL on a CpG (cytosine-phosphate-guanine) dinucleotide (cg18464137) located in the promoter of the *SEMA3E* gene (p-value = 9.73e-206; beta coefficient for C allele: 0.45), likely suggesting a silencing effect on the expression of the gene, which could be more prevalent in patients who did not show signs of disease activity according to our findings (OR_{EDA} for the C allele = 0.57; p-value = 2.15E-06). The Semaphorin-3E protein is part of the semaphorin superfamily, which is largely involved in the regulation of the immune system. The role of semaphorins has been the target for investigation in the past years. The semaphorins are mediators of the migration of oligodendrocyte precursors, potentially leading to recovery in demyelinating lesions of the CNS (Williams et al, 2007). In a recent work, the Semaphorin-4D was found to be the fundamental molecule mediating the microglia-astrocyte interaction underlying EAE, through its binding to the Plexin B1 and PlexinB2 (Clark et al, 2021). Semaphorin 3A and 4A expression is decreased in MS patients versus healthy controls and seem to have an opposite role in the activation of T regulatory cells (Eiza et al, 2022). Proteins of the semaphorin superfamily

have been found also in the whiter matter of MS lesions, supporting a role of Semaphorin 3A and 7A as mediator proteins in astrocytes and microglia/macrophages, that prevent remyelination in MS lesions (Costa et al, 2015).

The role of the Semaphorin 3E specifically, that we found potentially associated with disease activity in MS, has been less characterized by previous studies. Dendritic cells express *SEMA3E*, and its transcript binds to a specific receptor on the surface of NK cells and serves as a limiting factor for the migration of NK cells (Alamri et al, 2018). The Semaphorin 3E/Plexin D1 axis is an important therapeutic target in allergic asthma as well (Movassagh et al, 2019). *SEMA3E* is important in the proliferation and migration of Schwann cells, glial cells of the peripheral nervous system, and has been suggested as a potential target to enhance nerve regeneration (Shen et al, 2022). Given this body of evidence, *SEMA3E* represents a good candidate for future functional studies to assess the exact mechanism through which it could contribute to determine manifestations of disease activity in MS.

As mentioned, also the signal targeting the *NAMPT* gene looks promising in the context of MS. The lead variant in this locus is known to exert an eQTL effect in whole blood, and specifically on circulating monocytes (Fairfax et al, 2014). Moreover, it was found to have a mQTL effect in whole blood, with the minor allele T increasing the methylation level of cg05004518 in the gene body ($p=1.88e-89$; beta: -0.20; GoDMC, 2021), therefore likely leading to silencing of gene expression in immune cells (Weber et al, 2007). This finding could be of particular interest in the MS field, as inhibition of *NAMPT* has been shown to ameliorate Experimental Autoimmune Encephalitis (EAE) in mice (Meyer et al, 2022) and it has been proposed as a target to enhance neurological recovery in MS (Li et al, 2016).

NAMPT is the limiting enzyme in the conversion of nicotinamide to nicotinamide mononucleotide, which then results in the production of nicotinamide dinucleotide (NAD⁺), a metabolic process that is core in cellular redox reactions. Imbalance of NAD⁺ is reported in many diseases, including neurodegenerative diseases, aging, and immune diseases (Verdin E, 2015). Inhibition of the nicotinamide phosphoribosyltransferase, coded by the *NAMPT* gene, has been recently proposed as a therapeutic target in MS. *NAMPT* inhibition has been shown to limit the inflammatory properties of the astrocytes, through the means of CD38 and monocytes, resulting in an amelioration of the EAE in

mice (Meyer et al, 2022). We found that the rs2704966 T allele is associated to a lower probability of EDA and literature data support its role in down-regulating *NAMPT* in monocytes. This finding points towards the same direction of the work, showing that the anti-inflammatory activity of NAMPT inhibition goes through modulation of monocytes activity, that resulted in improved EAE outcome (Meyer et al, 2022). *NAMPT* inhibition has been also shown to mediate the neurological recovery in EAE mice treated with neural stem cells transfected with LINGO-1-Fc (Li et al, 2016).

Overall, it is interesting to note that both the above-mentioned mechanisms are in a close relationship with the function of oligodendrocytes and microglial cells, suggesting that they may contribute to an intrinsic regulatory balance of CNS inflammation. Perturbations to this fine mechanism, given by external triggers and influenced by underlying genetic variation, may therefore lead to clinical manifestations of disease activity, as explored by the NEDA-3 status.

These findings are additionally supported by our results on GO enrichment and pathway analysis, in which we found a significant enrichment for ‘*Response to interferon gamma*’ and ‘*Chemokine signaling*’ pathways, which survived adjustment for multiple testing. It is worth mentioning that most of the patients included in the study (~60%) started a treatment with interferon-beta at the baseline timepoint. With this in mind, such results appear to be even more biologically plausible. For example, the *JAK2* gene, which was part of the two above-mentioned pathways, is a key player in the activation of Th17 cells, the final effectors in the inflammatory cascade in MS (Conti et al, 2012).

As a limitation of our study, we recognize that the sample size is likely to be underpowered, as pointed out by the downward deflection of the curve in the QQ-plot (Figure 3.2). Limited statistical power does not allow to draw any final conclusions when exploring which are the genetic drivers of disease activity in MS. However, it is extremely difficult to find additional cohorts of patients that are very well characterized at the genetic level but also at the clinical and neuroradiological level, in order to be able to define the NEDA status for the enrolled patients. Nevertheless, our results can still at least partially address the question and hopefully prompt future investigations, paving the way to larger multicentric efforts.

4. Study 2: Epigenetic factors underlying disease activity

DNA methylation, the addition of a methyl group (-CH₃) at CpG dinucleotide sites, is one of the most stable epigenetic hallmarks across species (Mattei et al, 2022). Generally, methylation of CpG dinucleotides in enhancers and promoters leads to silencing of the genes, while methylation occurring in the gene body is usually linked to increased expression (Weber et al, 2007). The study of DNA methylation has been proven very effective in many areas of human biology and medicine, as it is a key element in the crosstalk between genetics, environment, and gene expression (Mattei et al, 2022). In medicine, as an example, the study of Differentially Methylated Positions (DMP) and Differentially Methylated Regions (DMR) has led to consistent advancements in understanding mechanisms of tumorigenesis and cancer transformation (Klutstein et al, 2016). A schematic representation of the difference between DMP and DMR is shown in Figure 4.1.

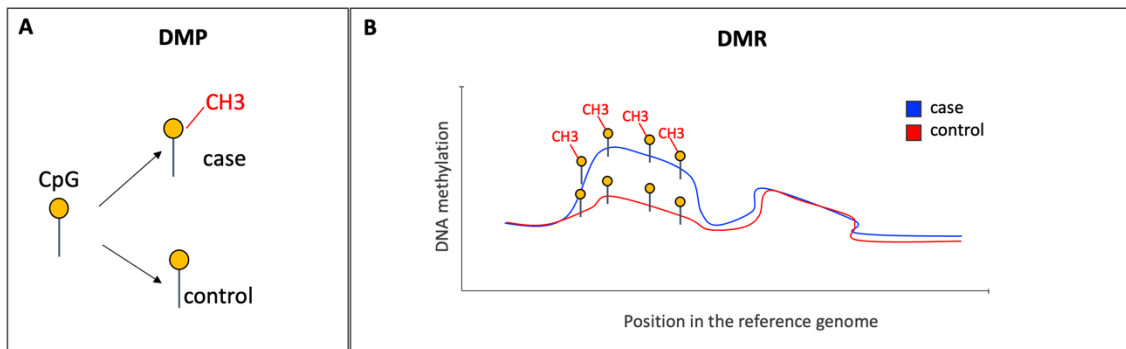


Figure 4.1. Schematic representation of Differentially Methylated Positions (DMP) and Differentially Methylated Regions (DMR). Each yellow pin represents a CpG dinucleotide site.

In the past few years, the study of epigenetics in general and of DNA methylation has greatly expanded in MS as well (Zheleznyakova et al, 2017). Many works have shown that changes in DNA methylation occur in MS patients compared to healthy controls, both in bulk tissue (Kulakova et al, 2016; Marabita et al, 2017) and isolated cell types (Baranzini et al, 2010; Maltby et al, 2015; Bos et al, 2016). DNA methylation is a recognized mediator of the risk of MS associated with the HLA-DRB1*15:01 allele (Kular et al, 2018). Conversely, very few studies have been published regarding the association of DNA methylation and measures of disease activity or severity in MS (Campagna et al, 2022).

In this work, we will explore the association between the methylation profile before the start of a first line DMT and disease activity assessed by NEDA-3 status after 2-years of follow-up. The study was performed in previously untreated MS patients from IRCCS San Raffaele Hospital (OSR) and at the Centre Hospitalier Universitaire de Toulouse (CHUT). Detailed inclusion criteria are reported in the Methods section (Chapter 9.2.1) Clinical features of the cohort at baseline are shown in Table 4.1.

	<i>OSR</i>	<i>CHUT</i>	<i>Total</i>	<i>p-value</i>
<i>N subjects</i>	75	174	249	-
<i>Age sampling</i>	37.1 (9.8)	40.1(10.8)	39.2 (10.3)	0.04
<i>AAO</i>	32.7 (8.9)	33.4 (10.2)	33.2 (10.1)	0.68
<i>F/M ratio</i>	1.59	3.97	2.89	0.0021
<i>Disease duration</i>	4.3 (6.5)	6.6 (7.7)	5.9 (7.2)	0.046
<i>EDSS at BL</i>	1.5 (1.0-2.0)	1.5 (1.0-2.5)	1.5 (1.0-2.0)	0.13
<i>EDA/NEDA 2 yr</i>	29/46	106/68	135/114	0.002
DMT started at BL				-
<i>DMF</i>	68%	42%	50%	
<i>Teriflunomide</i>	16%	34%	29%	
<i>Copaxone</i>	13%	8%	9%	
<i>Interferon</i>	3%	16%	12%	

Table 4.1. Baseline characteristics of the cohort. *N* = number; *AAO* = age at onset; *F/M ratio* = Female/Male ratio; *BL* = baseline. *DMT* = Disease-Modifying Treatment. In the rightmost column, the *p*-values from chi-square test or Mann-Whitney-Wilcoxon test are reported, to explore baseline differences between the two centers (*OSR* vs *CHUT*).

4.1 Results

4.1.1 Identification of Differentially Methylated Positions (DMP)

After performing rigorous QC as described in Chapter 9.2, we carried out a differential methylation analysis to unravel sites of the genome that are differentially methylated at baseline in patients who will have Evidence of Disease Activity (EDA) versus patients with NEDA. As a result, we found 7 DMP passing the multiple-testing correction. A list of the top 20 DMP and the mapped genes is reported in Table 4.2.

The top-associated CpG, cg27267436 ($p=6.4e-08$), maps to an intronic region of *APBA3* (Amyloid Beta Precursor Protein Binding Family A member 3) and it was found

<i>cpg</i>	<i>logFC</i>	<i>p-value</i>	<i>adj.Pval</i>	<i>CHR</i>	<i>BP</i>	<i>gene</i>	<i>genomic.loc</i>	<i>CpG island loc</i>
<i>cg27267436</i>	0.20	6.42E-08	0.046	19	3754012	APBA3	Body	island
<i>cg20025086</i>	-0.11	1.65E-07	0.046	12	109569130		IGR	opensea
<i>cg20308351</i>	-0.21	1.77E-07	0.046	7	3067980	CARD11	5'UTR	opensea
<i>cg22193657</i>	-0.09	2.61E-07	0.046	3	194948010	XXYLT1	Body	opensea
<i>cg19915997</i>	0.16	3.61E-07	0.046	3	15492725	COLQ	3'UTR	opensea
<i>cg25829490</i>	-0.18	4.00E-07	0.046	2	176988792	HOXD9	Body	shore
<i>cg12362502</i>	0.22	4.16E-07	0.046	6	43603544	MAD2L1BP	TSS200	island
<i>cg07146435</i>	-0.12	6.45E-07	0.061	10	100028499	LOXL4	TSS1500	island
<i>cg00352218</i>	-0.18	7.05E-07	0.061	6	19691654		IGR	shore
<i>cg07973246</i>	-0.17	1.20E-06	0.094	12	64238719	SRGAP1	Body	shore
<i>cg24764861</i>	0.17	1.34E-06	0.095	16	1495122	CCDC154	TSS1500	island
<i>cg11923320</i>	-0.13	1.47E-06	0.095	1	63783977		IGR	island
<i>cg14345857</i>	-0.16	1.62E-06	0.096	5	72742869	FOXD1	1stExon	shelf
<i>cg25911023</i>	0.14	1.73E-06	0.096	1	202776454	KDM5B	Body	island
<i>cg01144764</i>	0.26	2.34E-06	0.121	10	121633063	C10orf119	TSS1500	island
<i>cg24424115</i>	0.13	2.64E-06	0.129	3	58476822	KCTD6	TSS1500	shore
<i>cg18190847</i>	-0.17	4.60E-06	0.200	3	11195751	HRH1	5'UTR	shore
<i>cg06138439</i>	-0.09	4.62E-06	0.200	16	54973128		IGR	shore
<i>cg17886959</i>	0.16	5.58E-06	0.214	16	56642024	MT2A	TSS1500	shore
<i>cg02919861</i>	-0.20	5.80E-06	0.214	8	70090456		IGR	opensea

Table 4.2. Top 20 Differentially Methylated Positions (DMP) according to the NEDA/EDA status. *LogFC*= log fold-change in EDA vs NEDA; *adj.Pval* = 5% FDR adjusted *p*-value; *CHR* = chromosome; *BP*=base-pair position; *gene*=mapped gene according to University of California Santa Cruz (UCSC) RefSeq annotation; *genomic.loc*= location in the genome. IGR = intergenic region; TSS=transcriptional start site. CpG island loc = location related to UCSC CpG island.

to be hypermethylated in patients with EDA. Of note, APBA3 is an activator of the Hypoxia-Inducible Factor 1-alfa (HIF1A) gene in monocytes and macrophages at the sites of inflammation (Hara et al, 2017). HIF1A is a fundamental player in the inflammatory cascade in MS and it is extensively discussed in Chapter 7.

The cg20308351 ($p=1.77e-07$) maps to the *CARD11* (Caspase Recruitment Domain Family Member 11), that regulates NF-kB through BCL10 in B cells (Sommer et al, 2005) and a previous study has identified a DMR in *CARD11* in B lymphocytes of patients with MS (Maltby et al, 2018).

Among the top DMP, we identified the cg25829490, hypomethylated in EDA versus NEDA patients and mapping to the body of *HOXD9* (Homeobox D9), an element in the pathway of Transforming Growth Factor-Beta (TGF- β). In a previous study a region in the *HOXD9* gene was also found hypomethylated in the white matter of MS patients versus controls (Huynh et al, 2014).

4.1.2 Identification of Differentially Methylated Regions (DMR)

Starting from DMP, we also built DMR, continuous genomic regions that differ in methylation based on the phenotype (NEDA/EDA status, in our case)

In our analysis, conducted as described in the Methods (Chapter 9.2.4), we detected 4 DMR when comparing patients fulfilling the criteria for the NEDA-3 status at 2-years follow-up with patients who conversely showed signs of disease activity within 2-years from baseline (Table 4.3).

rank	chr	start	end	cpgs	maxdiff	meandiff	min.fdr	gene
1	19	2250901	2251067	4	-0.032	-0.028	6.12E-08	AMH
2	17	76037035	76037364	5	0.031	0.020	3.26E-08	TNRC6C
3	7	56515666	56516129	5	0.042	0.029	1.61E-08	LOC650226
4	19	22234980	22235850	8	0.067	0.022	6.76E-09	ZNF257

Table 4.3. Differentially Methylated Regions (DMR) in NEDA vs EDA patients. Rank = DMR rank using Fisher's combined probability method. Chr, start, end = genomic coordinates of the region (reference: GRCh37/hg19). Width = width of the region in base-pairs. CpGs=number of CpGs composing the region. Maxdiff, meandiff= maximum and minimum difference in methylation beta-value found in the region. Min.fdr = minimum smoothed FDR p-value. UCSC RefSeq annotated gene.

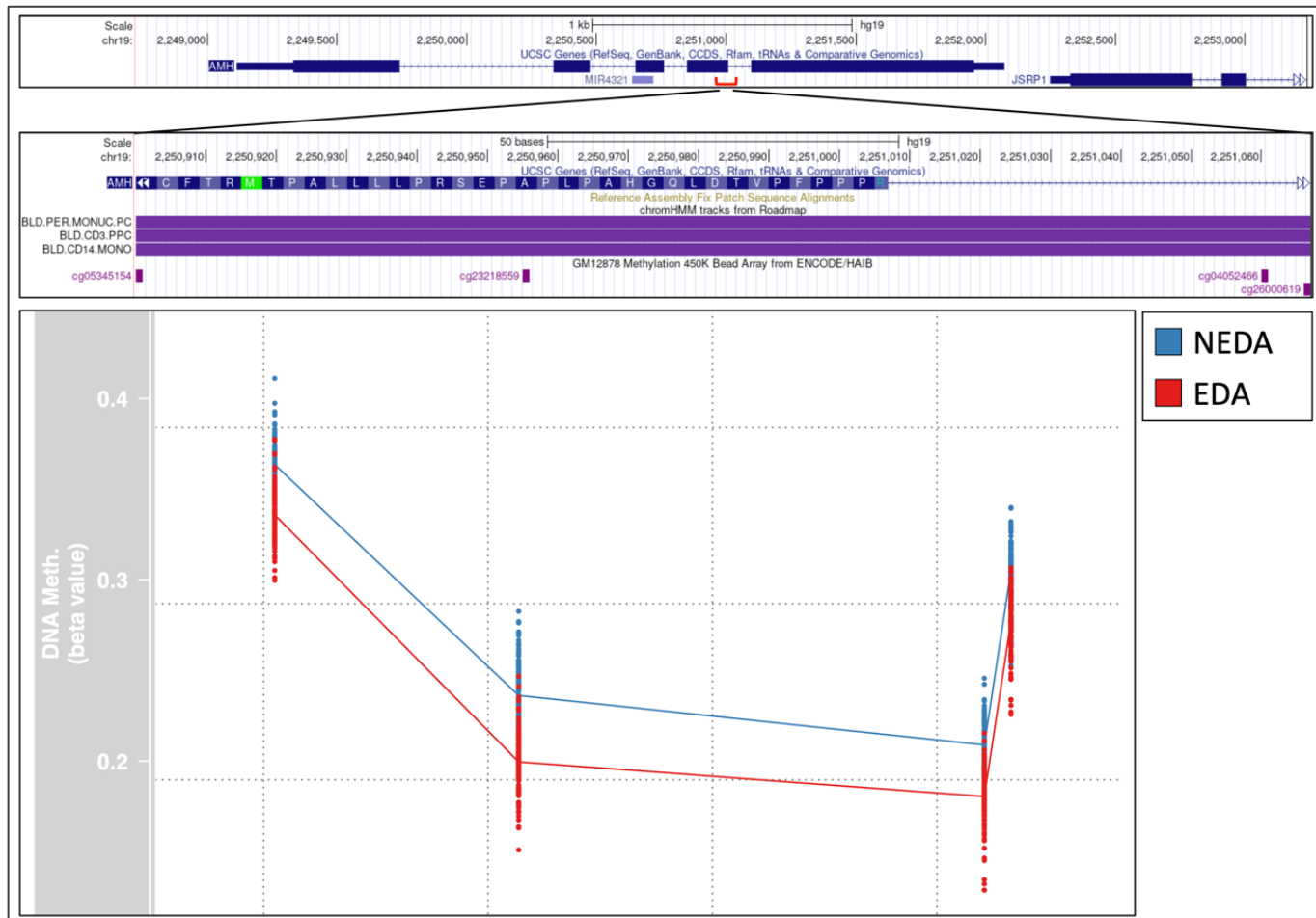


Figure 4.2. The identified Differentially Methylated Region in the AMH gene. On the top of the figure, the UCSC RefSeq track (GRCh37/hg19) for the AMH gene, with RoadMap chromatin states for PBMC, CD3+ and CD14+ cells. On the bottom, location of the DMR in the AMH gene (red bracket) and the visualization of the DNA methylation beta values according to the NEDA/EDA status.

In particular, after visual inspection and annotation of the DMR, the top-ranked region in chromosome 19 looked very interesting as it maps to the Anti-Mullerian Hormone (AMH) gene (Figure 4.2). The DMR is located in a 167 base-pairs wide region inside the bivalent promoter of AMH in the immune cells, as shown in Figure 4.2, therefore prompting a complex regulatory role in gene expression, as described in Chapter 4.2.

4.1.3 Methylation and expression Quantitative-Trait-Loci effect in AMH

To explore whether the difference in DNA methylation was driven by genetic factors, we extracted the SNPs mapping to a +/- 500 kilobase (kb) window from the DMR from our genome-wide meta-analysis on the NEDA-3 status in 1,408 patients described in Chapter 3. The top associated SNP mapping to the targeted region was rs2240656, located at ~135 kb from the DMR. In our analysis the rs2240656 C allele was associated with increased odds of reaching the NEDA-3 status at 2 years follow-up ($p=1.37e-03$, $OR_{NEDA}=1.73$, $I^2=0$).

In the attempt to explore whether the same variant could drive both the risk of disease activity and the difference in methylation in the *AMH* region, we assessed whether rs2240656 exerts a methylation Quantitative-Trait-Loci (mQTL) effect on the CpGs composing the DMR. Starting from the results of a mQTL study on whole blood in about 36,000 healthy subjects (GoDMC, 2021), we observed that the rs2240656 variant strongly increased the levels of methylation on all the 4 CpGs composing the DMR in the *AMH* gene. Then, we replicated this effect exploring the impact of rs2240656 on PBMC methylation in our cohort of RR-MS patients ($n=243$), finding strong evidence for a mQTL on the DMR (minimum p -value = $4.29e-10$) (Figure 4.3A). Results from mQTL effect calculation in PBMC of MS patients and comparison with data coming from literature are reported in Table 4.4.

CpG	β_{HC}	p_{HC}	β_{MS}	SE_{MS}	P_{MS}
cg26000619	0.71	0	0.081	0.012	4.29E-10
cg04052466	0.03	4.8E-03	0.066	0.066	1.75E-08
cg05345154	0.67	8.4E-290	0.073	0.073	3.00E-08
cg23218559	0.65	3.2E-269	0.079	0.013	1.67E-09

Table 4.4. Methylation Quantitative-Trait-Loci (mQTL) effect exerted by rs2240656 (C allele) on the 4 CpGs composing the identified DMR in AMH. To the left, the regression coefficient of the linear model (β_{HC}) and p-value in healthy controls (p_{HC}) from the GoDMC mQTL study in whole blood. To the right, the results of the mQTL effect analysis in PBMC from MS patients performed in this study, with the regression coefficient (β_{MS}), its standard error (SE_{MS}) and the p-value (P_{MS}).

In literature the rs2240656 C allele was also associated to a decreased expression of AMH in whole blood in data from a consortium (eQTLGen Consortium, 2018) studying the eQTL effect in ~31,000 subjects ($p=2.99e-20$). We were able to replicate this effect in PBMC from MS patients ($n=227$), as we found that the rs2240656 C allele was associated to decreased AMH expression ($p=7.41e-05$; β for C allele=-10.8) (Figure 4.3B).

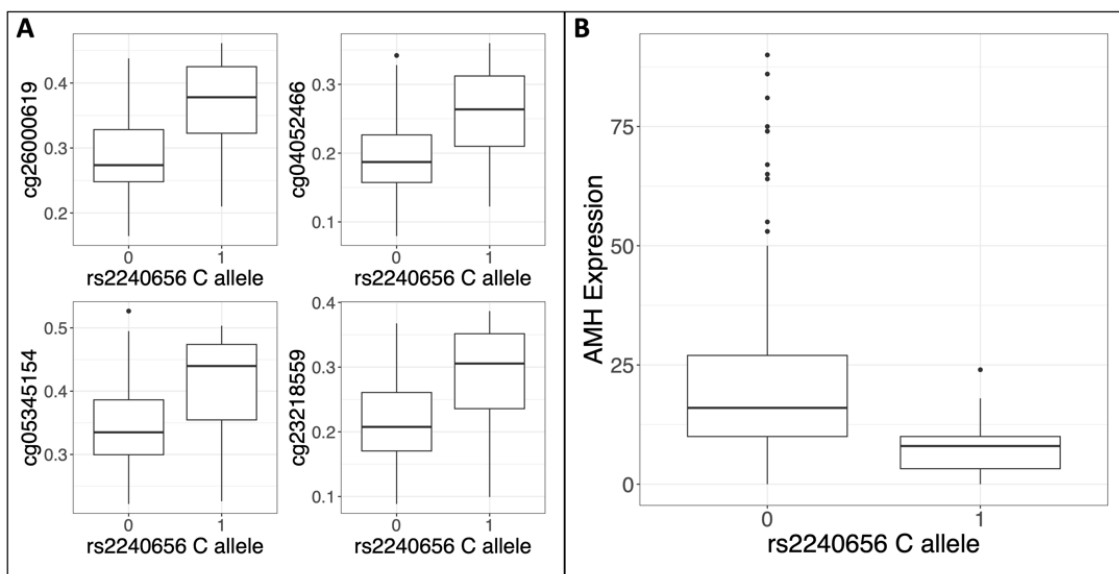


Figure 4.3. Effect of the rs2240656 variant on methylation and gene expression in MS. A: mQTL effect in PBMC from MS patients. y axis: methylation beta-values, x axis= number of rs2240656 C allele copies. B: eQTL effect in PBMC from MS patients; y axis: RNA-seq normalized counts, = number of rs2240656 C allele copies.

4.1.4 Sensitivity analysis

Interestingly, we rebuilt the DMR using the rs2240656 C allele status as a covariate and we found that the inclusion of the genetic factor prevented the detection of the DMR in *AMH*, additionally supporting that the effect on methylation is mainly driven by the genetic factor.

We also performed an additional sensitivity analysis to explore any bias related to the sex of the patients, even though in our initial analysis we had already used sex as a covariate for the identification of the DMP. First, *AMH* expression in PBMC was not different between male and female in our cohort of patients with MS, adjusting for age and center ($p=0.89$). Second, DMP and subsequent DMR analysis performed separately in males and females did not yield to the identification of the DMR in the *AMH* gene using the same cut-off when building the DMR (minimum smoothed FDR p -value $< 1e-07$), probably as a consequence of the reduced power, considering the smaller sample size of the studied cohorts. The DMR was identified in a secondary analysis in female patients only when considerably relaxing the threshold for statistical significance (minimum smoothed FDR- $p = 1.48E-05$), but not in male patients. This observation supports the hypothesis that statistical power prevents the identification when reducing the number of subjects, rather than a sex-specific effect.

4.1.5 Gene Ontology enrichment analysis

When running a GO enrichment analysis on the top 100 DMP, we found terms as '*Fertilization*' and '*Reproductive system development*', together with '*Leukocyte differentiation*', although the results were not significant at 5% FDR (Table 4.5).

4.2 Considerations

In this study, we analyzed whether CpG methylation in PBMC from MS patients before beginning a first line DMT was associated to the NEDA-3 status at 2-years. As mentioned, the studies investigating the role of methylation and disease activity in MS are very few. Our work led us to promising results, which can be divided in two different orders.

First, when investigating single CpGs, we found 7 DMP that were significant after FDR correction for multiple testing. Given their known functions, the most interesting

<i>GO term</i>	<i>Description</i>	<i>overlap</i>	<i>exp</i>	<i>ER</i>	<i>p-value</i>	<i>FDR</i>	<i>genes</i>
<i>GO:0001701</i>	in utero embryonic development	8/345	1.63	4.91	2.15E-04	0.18	APBA3;FOXD3;VEGFA;JAG2;ANKRD11;HES1;FZD5;HSD17B2
<i>GO:1903008</i>	organelle disassembly	4/96	0.45	8.81	1.10E-03	0.47	RNF41;MRRF;FZD5;STX5
<i>GO:0062012</i>	regulation of small molecule metabolic process	6/344	1.63	3.69	5.66E-03	1.00	HRH1;GUCA1B;ERLIN2;GCK;SQLE;SIRT5
<i>GO:0009791</i>	post-embryonic development	3/89	0.42	7.13	8.59E-03	1.00	KDM5B;VEGFA;FZD5
<i>GO:0002521</i>	leukocyte differentiation	7/496	2.34	2.99	8.81E-03	1.00	CARD11;STAT6;RNF41;VEGFA;JAG2;FZD5;DTX1
<i>GO:1902742</i>	apoptotic process involved in development	2/31	0.15	13.65	9.37E-03	1.00	JAG2;FZD5
<i>GO:0009566</i>	fertilization	4/176	0.83	4.81	9.62E-03	1.00	HOXD9;KDM5B;ATP8B3;TDRD12
<i>GO:0007219</i>	Notch signaling pathway	4/185	0.87	4.57	1.14E-02	1.00	JAG2;HES1;ANGPT4;DTX1
<i>GO:0001763</i>	morphogenesis of a branching structure	4/196	0.93	4.32	1.38E-02	1.00	FOXD1;KDM5B;VEGFA;FZD5
<i>GO:0061458</i>	reproductive system development	6/428	2.02	2.97	1.56E-02	1.00	KDM5B;VEGFA;ING2;HES1;FZD5;HSD17B2

Table 4.4. Results of Gene Ontology enrichment analysis. Size = size (number of genes) of the GO term. Overlap = number of genes in the GO term overlapping with the input genes. Exp = Expected enrichment. ER = enrichment ratio. P-value = nominal p-value for GO term. FDR: 5% FDR p-value. Genes = input genes overlapping with the GO term.

genes that emerged from this analysis are *APBA3*, *CARD11* and *HOXD9*. As already discussed, these genes have great importance in the context of MS, as they are known to be functionally linked to immune functions, and represent interesting targets to further explore their role in MS. The role of *APBA3* will be further discussed in Chapter 8. Of note, previous studies had already identified DMR in *CARD11* and *HOXD9* when comparing MS versus controls, therefore supporting the biological meaning of our findings, and prompting future functional studies as well as a replication in an independent cohort.

Second, when studying continuous regions of the epigenome, we detected 4 DMR. We specifically focused on the DMR on chromosome 19, mapping to the bivalent promoter of *AMH*. Bivalent promoters are typically found in genes involved in embryonic and fetal development. They present two kinds of histone modifications (H3K4me3 and H3K27me3) that allow a fast switch between silenced and activated state in specific cell types during development (Blanco et al, 2020). Indeed, during fetal life, *AMH* prevents the development of a female reproductive tract. Interestingly, AMH is part of the Transforming Growth Factor Beta (TGF- β) superfamily, a key player in the immune response (Johnston et al, 2016). In literature, despite many studies assessing the role of sex hormones in disease pathobiology, very little is known about AMH and MS (Ysrraelit MC and Correale J, 2018). In a first study, female patients with RR-MS were found to have lower AMH levels compared to healthy controls (Thöne et al, 2015). This finding was not confirmed by another study, in which AMH plasma levels were not different between RR-MS and controls, but patients with greater annualized relapse rate had significantly lower AMH levels, even though the sample size of the study was limited (Sepúlveda et al, 2016). In a larger longitudinal study, there was no difference in AMH levels between MS and healthy controls, but a decrease in AMH in plasma over time was associated with increased EDSS, worse MS Functional Composite (MSFC) score and greater grey matter atrophy in females with MS (Graves et al, 2018).

In the present study, we found that the rs2240656 C allele is associated with increased methylation in the identified DMR, decreased *AMH* expression in PBMC and whole blood, and reduced odds of having signs of disease activity. It is known that AMH levels reflect the aging of the reproductive system (Dewailly et al, 2014; Xu et al, 2019). In women, AMH decreases together with the number of follicles over the time and serves

also as a marker in conditions like the polycystic ovarian syndrome (Moolhuijsen et al, 2020). As the levels of AMH reflect aging, our observation, according to which patients who develop greater disease activity are genetically predisposed to a higher expression of *AMH*, looks reasonable. Indeed, younger patients typically have greater inflammatory disease activity, while with aging such risk decreases and the risk for progression increases (Confavreux et al, 2006). The above-mentioned study (Graves et al, 2018) found that AMH reduction was associated with measures of disease progression, as increased EDSS and grey matter atrophy, therefore supporting this hypothesis. It is also important to mention that while previous studies measured the concentration of the protein in blood, in our study involving the eQTL we measured the transcript (mRNA).

Of note, our work is the only one investigating the role of *AMH* in both men and women with MS. The AMH indeed has an important role in the reproductive function of male healthy subjects as well. The AMH is produced by the Sertoli cells in the testis during sex differentiation and it is suppressed by androgens when the primary differentiation is completed (Xu et al, 2019). At puberty, AMH levels rise again in the male and regulate fertility. As in women, AMH levels decrease over time in men (Xu et al, 2019). So, it is possible to hypothesize that an effect of aging on the reproductive system in general could be shared in women and men with MS. Interestingly, when we run a GO enrichment analysis on the DMP, we found an enrichment for many terms related with in utero development, fertilization, and embryogenic development, involving other genes (Table 4.4) and additionally supporting the hypothesis that sex hormones could be important in disease activity.

Further studies are needed to confidently establish the causal association between AMH and disease activity.

5. Study 3: Vitamin D and disease activity

Vitamin D is an important modulator of the immune system (Gombash et al, 2022). A neuroprotective effect of vitamin D has been also hypothesized, even though the exact mechanisms are still unclear. In the context of MS, higher vitamin D levels, through dietary intake or ultraviolet B radiation exposure, reduce the risk of developing the disease (Munger et al 2006; Munger et al, 2016).

When considering the relationship between vitamin D levels and measures of disease activity in MS, many questions are still open. Most of the published literature was focused on evaluating the adjunct benefit of vitamin D supplementation on disease-modifying treatment (DMT) (Ascherio et al, 2014; Stewart et al, 2012; Rotstein et al, 2015), with the results that were often contradictory, as regards for example treatment with interferon (Løken-Amsrud et al, 2012).

In this study, we aim to evaluate whether there is a causal association between vitamin D levels at baseline and disease activity levels measured by the NEDA-3 status at 2 years follow-up in a double-centric cohort of RR-MS patients sampled before the start of a first line DMT. The clinical features of this cohort are shown in Table 5.1 and the inclusion criteria are reported in Chapter 9.3.1.

	<i>OSR</i>	<i>CHUT</i>	<i>Total</i>	<i>p-value</i>
<i>N subjects</i>	65	165	230	-
<i>Age sampling</i>	36.6 (9.8)	39.7 (10.9)	38.8 (10.7)	0.051
<i>AAO</i>	32.7 (9.05)	33.5 (10.3)	33.3 (10.0)	0.58
<i>F/M ratio</i>	1.95	3.7	3.03	0.0021
<i>Disease duration</i>	3.9 (6.1)	6.2 (7.4)	5.5 (7.1)	0.035
<i>EDSS at BL</i>	1.5 (1.0-2.0)	1.5 (1.0-2.5)	1.0 (1.0-2.0)	0.29
<i>EDA/NEDA 2 yr</i>	28/37	95/70	123/107	0.046

Table 5.1. Baseline characteristics of the cohort. *N* = number; *AAO* = age at onset; *F/M ratio* = Female/Male ratio; *BL* = baseline. In the rightmost column, the *p*-values from chi-square test or Mann-Whitney-Wilcoxon test are reported, to explore baseline differences between the two centers (*OSR* vs *CHUT*)

5.1 Results

5.1.1 Vitamin D levels are associated with disease activity

After adjustment for vitamin D seasonal variation as described in Chapter 9.3 (see also Figure 9.3.1), we found that in our cohort of 230 RR-MS patients higher baseline levels of vitamin D were associated with higher probability of fulfilling the criteria for the NEDA-3 status after 2 years of follow-up ($p=0.019$) (Figure 5.2A).

Also, higher vitamin D levels at baseline were significantly associated with fewer relapses, as patients having 2 or more relapses during the follow-up had significantly lower vitamin D at baseline, if compared with patients having no relapses ($p<0.001$) and patients who experienced one relapse ($p=0.037$) (Figure 5.2B).

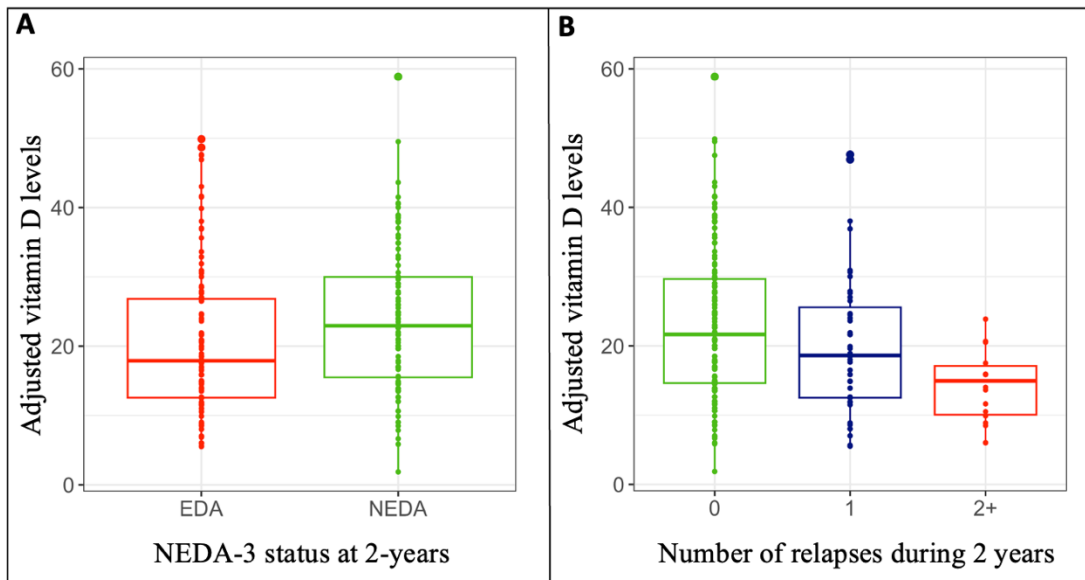


Figure 5.1. Disease activity at 2 years is affected by baseline vitamin D levels. *A:* Vitamin D levels at baseline and NEDA-3 status at 2-years. *B:* Vitamin D at baseline at number of relapses during the 2-years follow up (0, 1 and 2 or more). Vitamin D levels on the y axis are reported in ng/ml.

5.1.2 Genetic variation in vitamin D affects disease activity

We then investigated whether genetic variation in vitamin D levels was associated with disease activity, taking advantage of a previously published GWAS on 417,580 subjects that identified more than 140 genetic loci associated with vitamin D levels in the general population (Revez et al, 2020). First, we extracted from this study all the SNPs ($n=16,646$; no clumping for linkage disequilibrium) with genome-wide significant p -value ($<5e-08$).

Then, we assessed the association between these SNPs and the NEDA-3 status in our cohort, starting from the results of the meta-analysis that we conducted in Chapter 3, that involved a total of 1,408 subjects. In our meta-analysis 10,073 out of 16,664 SNPs were present. For some of these SNPs, we found nominal evidence for association with both NEDA-3 status and vitamin D levels (SNPs with meta-analysis p -value $< 0.01 = 143$) (Figure 5.2). A list of the top 10 associated loci with the lead SNP, the mapped genes, and the summary statistics of the GWAS on vitamin D levels (Revez et al, 2020) is reported in Table 5.2.

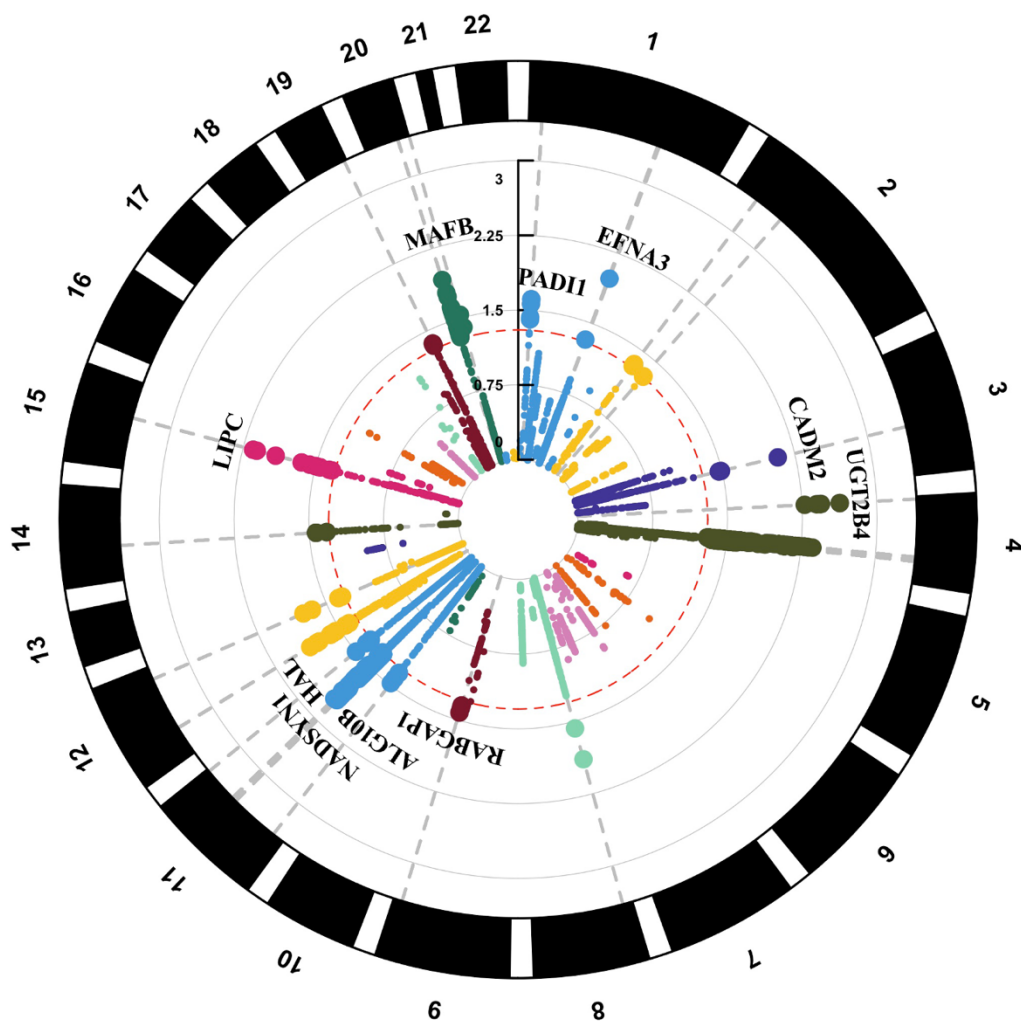


Figure 5.2. Circular Manhattan plot showing the association with NEDA-3 status for the SNPs with known role on vitamin D level variation in controls (Revez et al, 2020). The name of the gene is reported on top of the top 10 associated loci. Red dashed line = p -value < 0.05 . P -values are reported on $-\log_{10}$ scale.

<i>SNP</i>	<i>CHR</i>	<i>BP</i>	<i>AI</i>	<i>P_{NEDA}</i>	<i>OR_{NEDA}</i>	<i>Beta_{VitD}</i>	<i>P_{VitD}</i>	<i>Gene</i>
<i>rs62298881</i>	4	70008252	C	4.32E-03	1.44	0.024	1.13E-14	UGT2B4
<i>rs10468017</i>	15	58678512	C	7.10E-03	1.28	0.024	7.78E-30	LIPC
<i>rs9811546</i>	3	85391672	G	8.37E-03	0.79	-0.012	1.91E-08	CADM2
<i>rs3806256</i>	1	155035611	C	0.010	1.24	0.012	1.36E-09	EFNA3
<i>rs1790349</i>	11	71142350	T	0.011	1.35	0.30	6.97E-264	NADSYN1
<i>rs4812443</i>	20	39198836	T	0.012	0.80	-0.012	1.74E-08	MAFB
<i>rs78633929</i>	12	96278208	T	0.018	1.29	0.014	4.19e-08	HAL
<i>rs1608906</i>	12	33894986	T	0.017	0.82	-0.011	2.38E-08	ALG10B
<i>rs34130414</i>	1	17560262	G	0.024	1.21	0.020	2.27E-22	PADI1
<i>rs12351386</i>	9	125855711	C	0.038	0.75	-0.018	1.75E-08	RABGAP1

Table 5.2. Genetic loci that are associated with serum vitamin D levels variation in controls and with NEDA-3 status. $Beta_{VitD}$ and P_{VitD} refer to the regression coefficient and the p-value from the linear model investigating the influence of the SNP on vitamin D levels from Revez et al, 2020. P_{NEDA} and OR_{NEDA} refer to the GWAS meta-analysis on NEDA-3 status described in Chapter 3. Gene = nearest gene according (GRCh37 coordinates).

5.1.3 Mendelian Randomization analysis

Then, to assess whether the observed association between vitamin D levels is causal and it is not confounded by other factors, we performed a Mendelian Randomization (MR) analysis taking advantage of the above-mentioned GWAS (Revez et al, 2020). The results of our Inverse Variance Weighted (IVW) analysis supported the presence of a causal effect ($p=7.79e-06$), as the probability of having no disease activity over a 2-year follow-up increases with genetically predicted increase of vitamin D levels. Additional methods also supported this finding, as shown in Table 5.3 and Figure 5.4. Detailed methodological information regarding the MR analysis is reported in the Methods section (Chapter 9.3.3)

<i>Method</i>	<i>N IVs</i>	<i>Beta</i>	<i>SE</i>	<i>p-value</i>
<i>Inverse variance weighted</i>	15	0.085	0.019	7.79e-06
<i>Weighted median</i>	15	0.057	0.012	8.524e-07
<i>MR Egger</i>	15	0.16	0.087	0.087
<i>Weighted mode</i>	15	0.056	0.013	8.08e-04
<i>Weighted mode (NOME)</i>	15	0.057	0.0046	6.36e-09
<i>Simple mode (NOME)</i>	15	0.059	0.0046	4.26e-09

Table 5.3. Results of the Inverse variance weighted Mendelian Randomization analysis and comparison with different methods. *N IVs* = number of instrumental variables. *Beta* = regression coefficient and its standard error (*SE*). *NOME* = No Measurement Error in the SNP Effect.

5.2 Considerations

Vitamin D is a known risk factor for MS (Olsson et al, 2017). Since vitamin D has an effect on the primary events that lead to MS, it is reasonable to hypothesize that it may also play a role in determining disease manifestations once the biological cascade of MS has already been triggered, and therefore impacting on disease activity.

From this hypothesis, in this study we investigated the association of vitamin D with the NEDA-3 status, assessed at 2-years from the beginning of a first line DMT. We found that higher vitamin D levels are associated with lower probability of having disease manifestations assessed by the NEDA-3 status and lower number of relapses.

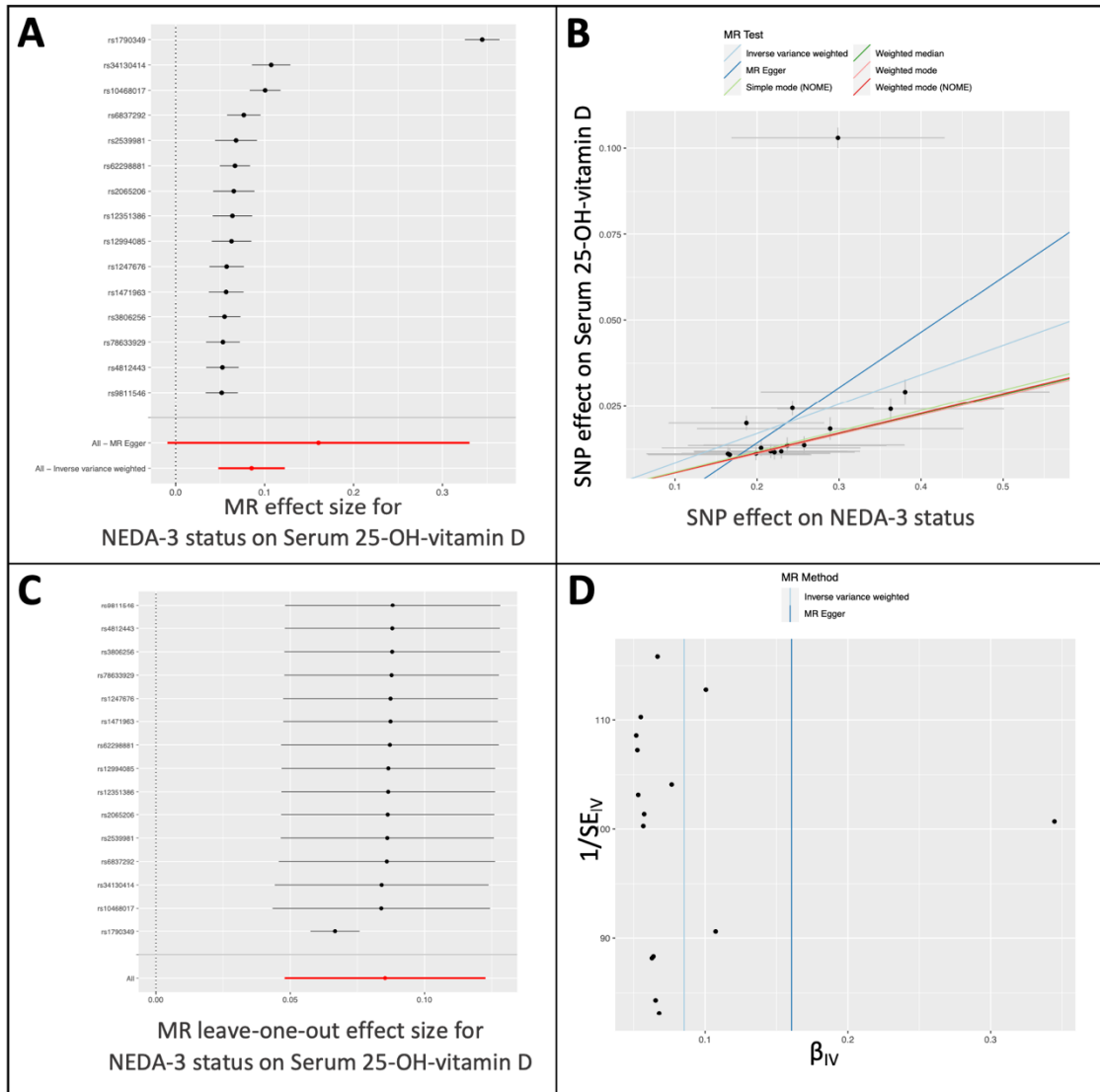


Figure 5.3. Mendelian Randomization analysis of vitamin D and NEDA-3 status. Figure A: forest plot showing the effect of the single IVs and the cumulative effect for the MR Egger and the Inverse-variance weighted MR analysis. The plot shows the regression coefficient (beta) and 95% confidence interval of the odds of fulfilling the criteria for the NEDA-3 status per a 1-standard deviation increase in genetically predicted vitamin D levels. Figure B: Comparison of different MR methods, in which the direction of the effect is the same. Figure C: leave-one-out sensitivity analysis, confirming positive correlation between genetically predicted vitamin D levels and NEDA-3 status, despite removing the IV with the biggest effect. Figure D: funnel plot showing causal estimation (beta coefficient for the IVs on x-axis) and the strength of the instrument variables ($1/SE = 1/\text{standard error}$) on the y-axis.

Few works have explored the association between vitamin D and disease activity. The results of observational studies on disease activity have been less clear if compared to what has been published regarding vitamin D as a risk factor for the disease, and likely suffered from small sample size and confounding factors (Smolders et al, 2019). In a

prospective observational study involving 1,482 RR-MS patients treated with interferon, lower vitamin D levels were associated with measures of disease activity at MRI, but not with change in the EDSS score, brain volume loss or relapse risk (Fitzgerald et al, 2015). In an observational study investigating both RR-MS and primary progressive MS, lower vitamin D levels at baseline were associated with greater probability of gadolinium enhancing lesions at MRI during a 2-year follow up, only in the relapsing-remitting group (Cree BA et al, 2016). Along with these and other observational studies (Smolders et al, 2019), the first randomized trials on vitamin D supplementation in association with DMT in RR-MS were published, assessing different outcomes, including conversion to RR-MS in patients with optic neuritis, MRI activity in RR-MS, EDSS change or risk of relapses. The results were often controversial, given the small sample size and the presence of many potential confounding factors (Smolders et al, 2019).

To date, to the best of our knowledge, only one study has involved the NEDA-3 status as an outcome for a randomized trial on vitamin D in MS (Hupperts et al, 2019). In this randomized trial involving 229 patients with RR-MS treated with interferon beta-1a and vitamin D supplementation or interferon beta-1a and placebo, there was no difference in the probability of reaching a NEDA-3 status at 48 weeks, but patients with higher vitamin D had better outcome when examining MRI measures of disease activity (Hupperts et al, 2019). In this study, the short follow-up time (48 weeks) may have limited the number of events needed to observe a significant effect in a small cohort.

Interestingly, in a genetic study assessing the relapse hazard in untreated MS patients, the authors found a significant enrichment for genes belonging to the vitamin D pathway (Vandebergh et al, 2021), and a subsequent MR analysis was successful in determining the causal role of vitamin D on relapse hazard (Vandebergh et al, 2022).

Indeed, MR is a very powerful tool to assess causal inference when high-quality randomized controlled trials are missing or in presence of many potential confounding factors. By the use of genetic markers (which are therefore named as instrumental variables, IVs), which by definition are distributed in a random way in the population and serve as a randomization strategy, MR allows to establish whether an exposure (e.g. vitamin D) is causally associated with an outcome (= disease activity). Therefore, in our study, after finding that vitamin D levels at baseline predict the NEDA-3 status at 2-years,

we tried to gain a deeper insight into this association, investigating whether there is a causal relationship, implementing a MR analysis.

Our results overall provide evidence that not only vitamin D levels are associated with the NEDA-3 status at 2-years, a powerful outcome of disease activity which was poorly explored by previous studies on vitamin D, but also that there is a causal relationship in this effect which does not seem affected by other potential confounding factors and that is likely due to the many known effects of vitamin D in the immune system (Gombash et al, 2022).

We reckon that the evidence provided by our work can promote the need for future further randomized trials assessing the effect of vitamin D supplementation on outcomes of disease activity, as the NEDA-3 status, in a real-world setting.

6. Study 4: Environmental factors and disease severity

The pathogenesis of MS is complex and multi-factorial. The study of environmental factors (EF) is one of the main candidates to successfully examine the mechanisms involved in the disease and to fill the gap left by the genetic studies. As more extensively discussed in Chapter 1, many EF have been showed to impact the susceptibility to develop the disease (Olsson et al, 2017), but much less is known on their effect on measures of disease severity in MS (Healy et al, 2009; Manouchehrinia et al, 2013; Ramanujam et al, 2015). Previous evidence also has suggested that the role of the EF could be stronger during an early timeframe in the disease course, representing the scenario in which, on the basis of the underlying genetic background, the pathogenic mechanisms of MS are triggered (Kular and Jagodic, 2020).

Starting from this hypothesis, in this study we will investigate how the exposure to specific EF occurred early in the disease history of MS patients, that is during adolescence and before the diagnosis of MS, can potentially affect future MS severity, assessed by the Age-Related Multiple Sclerosis Severity (ARMSS) score.

Taking advantage of an extensive environmental questionnaire (EQ) that was administered to a large cohort of consecutive MS patients at the MS Centre of IRCCS San Raffaele Hospital (n=1,892), we will focus on: a) factors provoking changes in sex hormones in women with MS; b) body weight; c) alcohol consumption.

The clinical characteristics of the final study cohort, that included 1,688 patients, are reported in Table 6.1. Further details on patient selection, EQ and statistical analysis are described in the Methods section (Chapter 9.4).

<i>Number of patients</i>	<i>1,688</i>
<i>F/M ratio</i>	1.96 (F: 1,117; M: 571)
<i>Age at EQ (mean, sd)</i>	51.1 (8.6)
<i>Age at onset (mean, sd)</i>	31.1 (9.5)
<i>Age at ARMSS (mean, sd)</i>	52.4 (6.1)
<i>EDSS at ARMSS (mean, sd)</i>	2.3 (1.8)
<i>Disease duration at ARMSS (mean, sd)</i>	15.1 (9.4)

Table 6.1. Clinical features of the cohort included in the study on environmental factors. F/M = female/male. EQ = environmental questionnaire; sd = standard deviation. EDSS = Expanded Disability Status Scale score.

6.1 Results

6.1.1 Factors provoking changes in sex hormones

In the subset of female patients (n=1,117), we explored the role of hormonal factors on disease severity in MS. In this regard, the EQ assessed the following topics: the age at menarche, the use of oral contraceptives, history of pregnancy before the diagnosis and occurrence of menopause before the diagnosis.

We found a trend for association between the age at menarche and the ARMSS score, with patients who reported an age at menarche ≥ 14 years (n=264) having a less severe disease (p=0.076). Interestingly, when assessing the use of oral contraceptive before the diagnosis, we found that patients who reported to have used oral contraceptives prior to the diagnosis had a better outcome in terms of future MS severity (p=0.0048). Of note, also patients who reported a pregnancy before the diagnosis had a lower ARMSS score (p=0.0015), as well as patients in whom menopause (n=60) had occurred before the diagnosis (p<0.001). Assessing whether the use of oral contraceptive or pregnancy could have influenced the age at onset in MS (which is known to be associated with disease activity and severity), only the history of pregnancy resulted to be associated (p<0.001) and not the use of oral contraceptive (p=0.56) or the age at menarche (p=0.11).

Conversely, the use of oral contraceptive at diagnosis was not associated with disease severity (p=0.182).

The results of these analysis are shown in Table 6.2 and Figure 6.1.

<i>Factor</i>	<i>Beta [SE]</i>	<i>p-value</i>	<i>Response rate</i>
<i>Age at menarche ≥ 14 years</i>	-0.13 [0.071]	0.076	99%
<i>Oral contraceptive before diagnosis</i>	-0.17 [0.061]	0.0048	98%
<i>Pregnancy before diagnosis</i>	-0.20 [0.064]	0.0015	98%
<i>Menopause before diagnosis</i>	-0.50 [0.13]	< 0.001	98%
<i>Contraceptive use at diagnosis</i>	0.095 [0.071]	0.182	98%

Table 6.2. The impact of age at menarche, oral contraceptive use, menopause, and pregnancy on disease severity. P-value, regression coefficient (beta) and its standard error (SE) of linear models using normalized ARMSS score as outcome are reported. Response rate = percentage of patients answering the question in the EQ.

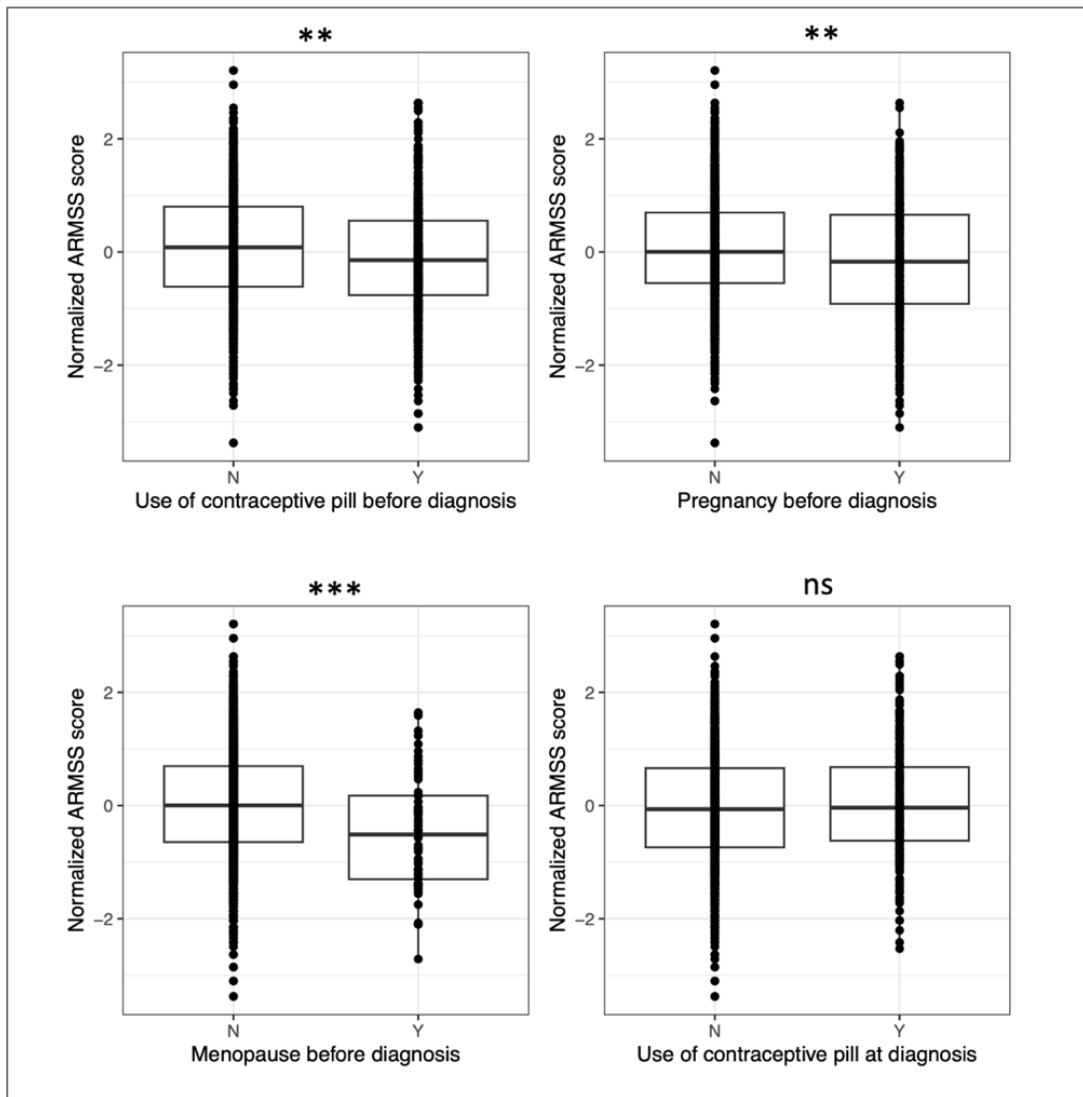


Figure 6.1. Box-dot plot showing the association of oral contraceptive use, menopause and pregnancy with the ARMSS score. Significance threshold: ** = $p < 0.01$; *** = $p < 0.001$; ns = not significant. N = No; Y = Yes

6.1.2 Body weight

When we evaluated the body weight during adolescence and disease severity, we did not find any significant correlation. In particular, patients who self-reported that they were obese ($n=19$; $p=0.10$), overweight (305 ; $p=0.16$), or underweight (184 ; $p=0.28$) during the adolescence did not have a different disease course in terms of ARMSS score if compared to patients who reported normal body weight during adolescence ($n=1,180$). Even after grouping together patients who were obese or overweight (given the small

number of patients who reported adolescence obesity), there was no significant effect on disease severity ($p=0.11$).

When we evaluated how the Body Mass Index (BMI) at the time of diagnosis impacted future disease severity, we found that patients with higher BMI had a smaller risk of a more severe course assessed by the ARMSS score ($p=0.018$; Figure 6.2). Grouping the BMI into different categories (underweight, normal, overweight, obese), we found only a trend for association for underweight patients, when comparing each category to the normal weight category as reference ($p=0.062$). In a sensitivity analysis, we found that patients who were overweight or obese at the time of diagnosis, had a higher age at disease onset ($p<0.001$). When adjusting the analysis of BMI at diagnosis and future severity at ARMSS score for the age at onset, the previously reported association did not hold ($p=0.58$). The results of the above ported analysis are shown in Table 6.3.

<i>Factor</i>	<i>Beta [SE]</i>	<i>p-value</i>	<i>Response rate</i>
<i>Weight adolescence:</i>			100%
<i>underweight vs normal</i>	0.084 [0.08]	0.28	
<i>overweight vs normal</i>	0.091 [0.064]	0.16	
<i>obese vs normal</i>	0.38 [0.23]	0.10	
<i>obese/overweight vs normal</i>	0.38 [0.23]	0.11	
<i>Weight at diagnosis</i>			96%
<i>BMI</i>	-0.058 [0.025]	0.018	
<i>underweight vs normal</i>	0.18 [0.95]	0.062	
<i>overweight vs normal</i>	-0.074 [0.058]	0.20	
<i>obese vs normal</i>	0.020 [0.11]	0.85	

Table 6.3. Body weight and disease severity assessed by the ARMSS score. P-value, regression coefficient (beta) and its standard error (SE) of linear models using normalized ARMSS score as outcome are reported. Response rate = percentage of patients answering the question in the EQ.

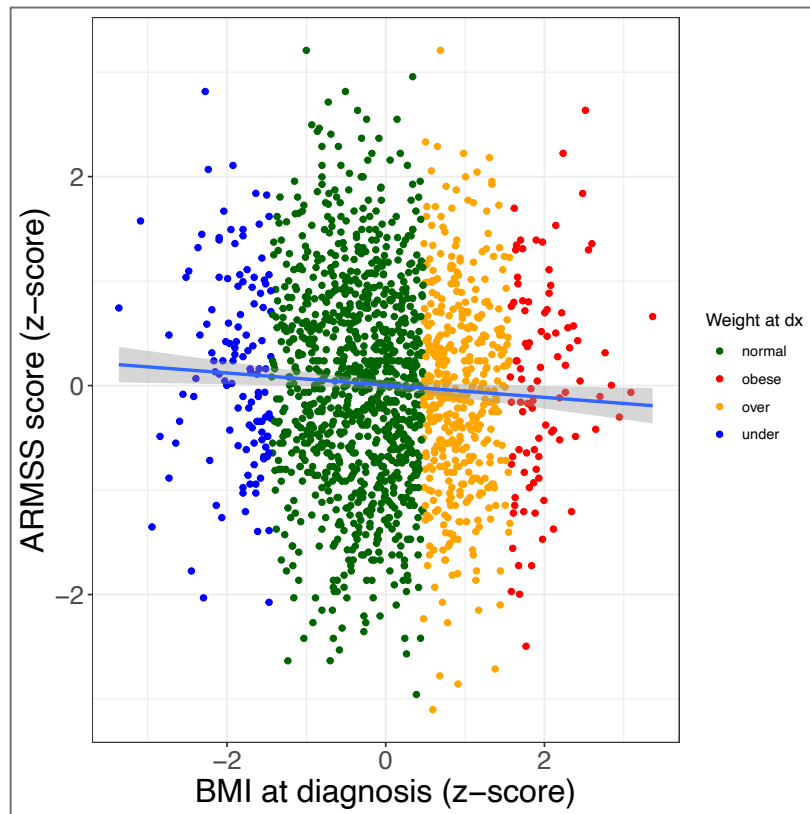


Figure 6.2. Impact of BMI at the time of diagnosis on future disease severity. BMI score is shown in a form of a z-score to fit a normal distribution, as well as for the ARMSS score. The blue line represents the smooth line of the linear model and its 95% confidence interval (grey area). Categories for weight at diagnosis (dx) are defined as: normal (BMI 18-24), under (BMI < 18), overweight (BMI 24-30), obese (BMI > 30).

6.1.3 Alcohol consumption

Alcohol consumption was studied separately in women and men, due to the known gender differences in the metabolism of alcohol (Baraona et al, 2001). All the analyses regarding alcohol consumption are adjusted for smoking status and body weight, as suggested considering that both affect alcohol intake and metabolism (Hedström et al, 2014b; Olsson et al, 2017). Details on the classification of patients into categories of alcohol consumption (no/moderate/high) are discussed in the Methods (Chapter 9.4.4).

6.1.3.4 Alcohol consumption in women

We found that women who reported alcohol consumption during adolescence had lower disease severity explored by the ARMSS score ($p=0.0030$). When dividing patients based on the level of consumption, this finding was confirmed comparing women who reported a moderate consumption ($n=360$) versus women who reported no alcohol

consumption (n=748) during adolescence (p=0.0038) (Figure 6.3A). No association was found in women who reported a high consumption versus no consumption during adolescence, probably due to a more limited number of patients reporting a high consumption during adolescence (n=9).

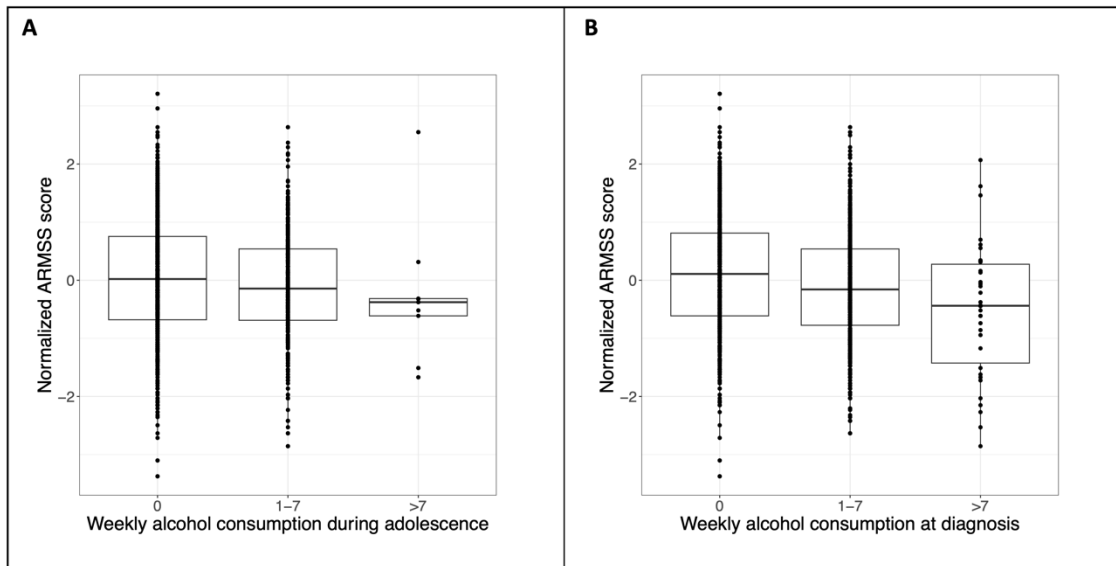


Figure 6.3. Alcohol consumption and MS severity in women.

The same effect was found also when investigating alcohol consumption at the time of MS diagnosis. In particular, women who reported alcohol consumption at diagnosis had significantly less severe MS during disease course (p=7.69e-07), and the same effect was found for all the levels of alcohol consumption when dividing into categories (No versus Moderate: 4.53e-06; No versus High: 0.00064) as show in Figure 6.3B.

In the subset of patients who reported drinking alcohol, we also investigated whether the kind of alcoholic beverage was associated with severity measure. Interestingly, patients who reported consumption of beer or wine seemed to benefit from a protective effect towards disease severity (p=0.019), while consumption of spirits showed a trend for association with worse disease severity (p=0.067). We did not find the same protective effect for beer and wine when analyzing reported alcohol consumption during adolescence. The above-mentioned results are reported in Table 6.5.

Female patients with MS (n=1,117)

<i>Factor</i>	Beta [SE]	p-value	Response rate
<i>Weekly alcohol at diagnosis</i>			100%
<i>No vs Yes</i>	-0.32 [0.06]	7.69e-07	
<i>No vs Moderate</i>	-0.30 [0.065]	4.53e-06	
<i>No vs High</i>	-0.65 [0.19]	0.00064	
<i>Spirits at diagnosis</i>	0.22 [0.12]	0.067	100%
<i>Beer/wine at diagnosis</i>	-0.35 [0.15]	0.019	100%
<i>Weekly alcohol adolescence</i>			100%
<i>Yes vs No</i>	-0.19 [0.064]	0.0030	
<i>No vs Moderate</i>	-0.19 [0.065]	0.0038	
<i>No vs High</i>	-0.31 [0.34]	0.36	
<i>Spirits during adolescence</i>	-0.036 [0.11]	0.73	100%
<i>Beer/wine during adolescence</i>	0.0040 [0.12]	0.97	100%

Table 6.5. Alcohol consumption in women and disease severity. The results of linear models using the ARMSS as outcome are shown, with p-value, regression coefficient (beta) and its standard error (SE).

6.1.3.5 Alcohol consumption in men

Investigating alcohol consumption at the time of diagnosis, our study yielded similar results, despite being more limited by statistical power considered the lower incidence of MS in men (n=571). Higher alcohol consumption at diagnosis was associated with a protective effect towards disease severity (p=0.031), but we did not find any significant association exploring adolescence (Figure 6.4).

As regards the kind of alcoholic beverage, we did not find any significant association. The results of the analysis in males are reported in Table 6.6

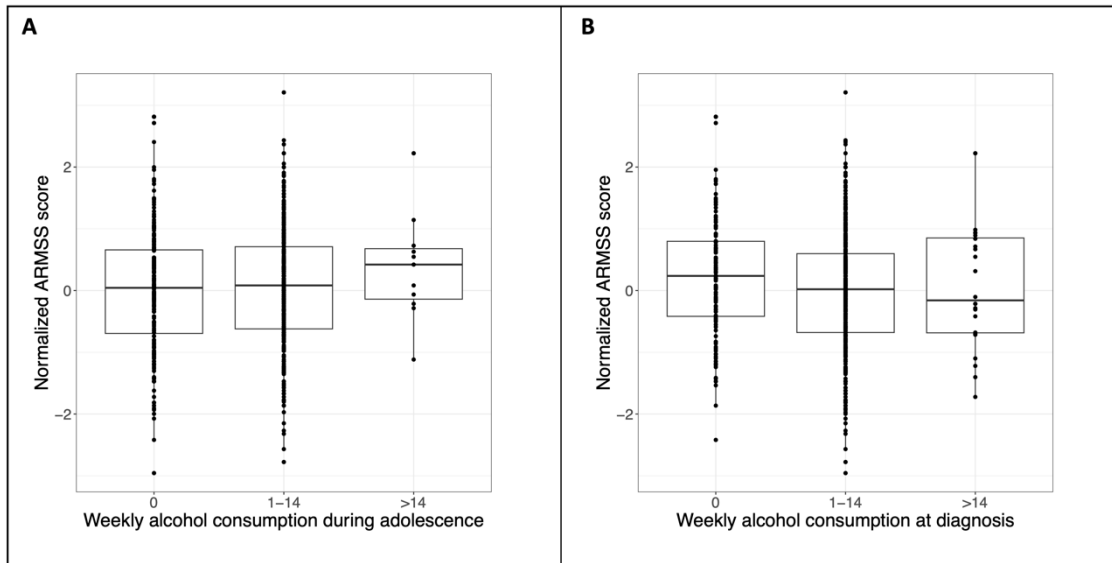


Figure 6.4. Alcohol consumption and disease severity in men.

Male patients with MS (n=571)

<i>Factor</i>	Beta [SE]	p-value	Response rate
<i>Weekly alcohol dx</i>			100%
<i>No vs Yes</i>	-0.23 [0.11]	0.031	
<i>No vs Moderate</i>	-0.24 [0.11]	0.025	
<i>No vs High</i>	-0.039 [0.23]	0.86	
<i>Spirits at dx</i>	0.16 [0.11]	0.16	100%
<i>Beer/wine at dx</i>	0.031 [0.19]	0.87	100%
<i>Weekly alcohol adolescence</i>			100%
<i>Yes vs No</i>	0.060 [0.084]	0.48	
<i>No vs Moderate</i>	0.051 [0.085]	0.55	
<i>No vs High</i>	0.36 [0.30]	0.23	
<i>Spirits during adolescence</i>	-0.026 [0.10]	0.80	100%
<i>Beer/wine during adolescence</i>	0.024 [0.082]	0.77	100%

Table 6.6. Alcohol consumption in men and disease severity. The results of linear models using the ARMSS as outcome are shown, with p-value, regression coefficient (beta) and its standard error (SE). Dx = diagnosis.

6.2 Considerations

6.2.1 Sex hormones and disease severity

Gender difference significantly impacts MS, which is more common in women but, once established, the disease course seems to be less favorable in men. The biological mechanisms through which sex hormones can affect the immune system are still not completely understood. Estrogens are able to increase the T regulatory lymphocytes component, down-regulating the effector Th17 lymphocytes in preclinical models (Polanczyk et al, 2004; Wang et al, 2009). Studies investigating a possible role of sex steroids on MS severity are few (Bove and Chitnis, 2014). In a small study on 132 patients, the use of oral contraceptive was associated with more favorable MS course, assessed by EDSS change and Multiple Sclerosis Severity Score (MSSS) (Sena et al, 2012). These results were replicated in another independent study on 174 women with MS, in which the use of oral contraceptive was associated with less severe disease course and resulted to be protective towards disease progression (Gava et al, 2014). In another study, RR-MS patients with at least two pregnancies had reduced risk of reaching an EDSS 6.0, while in PP-MS patients in the same study oral contraceptive use significantly correlated with a worse EDSS (D'hooghe et al, 2012).

Our study implied a large cohort of 1,117 women with MS, who showed a more favorable disease outcome if they reported use of oral contraceptive before the diagnosis. When we studied oral contraceptive at the time of the diagnosis and future MS severity, we did not find any significant association, additionally supporting that the role of such factors could be important in an early stage of the disease, before the pathogenic cascade of MS is fully established.

Overall, our findings replicate the above-mentioned evidence coming from smaller studies and prompts future prospective investigations.

6.2.2 Body weight and disease severity

As regards the effect of body weight, our results do not provide such a level of confidence. In literature, solid evidence demonstrated that obesity during adolescence leads to a higher risk of developing MS (Munger et al, 2012; Wesnes et al, 2015), interacting with the HLA risk genes (Hedström et al, 2014a). Adolescence is thought to be the critical period in which an increased body weight can affect the risk of the disease,

as in patients with a high BMI during adulthood or childhood there was no effect on the risk of MS (Hedström et al, 2016). Studies investigating the effect of body weight on MS severity are few. In a cohort of 351 people with MS, obese and overweight patients had a worse MSSS (Van Hijfte et al, 2022). In our study, we did not find any impact of obesity during adolescence on future MS severity. This result could be affected by the recall bias, as patients were asked to independently report their body size during adolescence. Prospective studies systematically assessing the BMI in children, adolescents, and young adults, despite having the disadvantage of observing a much more limited number of cases (limited by the incidence of MS), are needed to assess this effect more precisely. The recall bias may also have limited our findings in terms of power, as only 19 patients reported that they were affected by obesity in their adolescence. When assessing the BMI at diagnosis, instead, we found that an increase in BMI was linearly associated with a reduced ARMSS score, suggesting that patients with increased body weight at the time of the diagnosis were less prone to develop severe MS. When adjusting this analysis for the age at disease onset, we were not able to observe this correlation anymore, probably reflecting a bias due to patients' age, as older patients tend to have a higher BMI and less severe MS.

6.2.3 Alcohol consumption and disease severity

In the analysis on alcohol, we found that alcohol consumption was associated with less severe disease in women in a dose dependent manner, both when investigating alcohol consumption during adolescence or at the time of diagnosis. In the subset of men, we found the same effect only when assessing alcohol consumption at the time of diagnosis, while no association was found for adolescence, probably due to a much more limited sample size if compared to women. Overall, alcohol exerts an inhibitor effect on the immune system, targeting both innate and adaptive immunity (Fahim et al, 2020). Alcohol is able to down-regulate the activity of T lymphocytes, limiting the function of the antigen-presenting cells as monocytes and dendritic cells (Heinz and Waltenbaugh, 2007; Ness et al, 2008; Sureshchandra et al, 2019). It is known to induce wide changes in the cytokine asset, for example repressing IL-6 and TNF-alfa (pro-inflammatory factors) and increasing the levels of IL-10 (that has anti-inflammatory activity) (Sureshchandra et al, 2019). Together with the underlying biological explanation, the positive effect of

alcohol consumption on MS has been reported by some other studies, both on the susceptibility to the disease and its severity (Hedström et al, 2014b; Diaz-Cruz et al, 2017). Similar evidence has been found on other autoimmune diseases like rheumatoid arthritis (Källberg et al, 2009). Overall, the epidemiological evidence is controversial, as others found no effect by alcohol consumption on disease severity, but its role on attenuating the risk given by smoking, which is largely known to worsen the outcome of MS, was postulated (Ivashynka et al, 2019). Since many people who drink alcohol are also smokers, in our analysis we tried to minimize the bias of smoking as a confounding factor adjusting for smoking history, as also done by others (Hedström et al, 2014b), probably picking up a true effect of alcohol on severity. In this regard, our study provides a validation of previous studies in a large cohort of MS patients, both women and men.

Interestingly, in the analysis on women, we also found that the kind of alcohol beverage that patients reported to drink had an opposite effect on severity, as wine and beer were associated with better outcome, while patients reporting consumption of spirits had higher ARMSS score. This finding, even though a validation in independent and prospective cohorts is strongly needed, is potentially of great interest, as a neuroprotective effect of additional compounds that are present in some kind of alcohol beverages is well known. For example, a consistent body of evidence has shown that resveratrol and other polyphenols, which are especially found in red wine, have known antioxidant properties, exerting a neuroprotective effect which is associated with a decrease in the aging process (Zhou et al, 2021).

6.2.4 General considerations

Overall, in this study we provide evidence supporting that exposure to some kind of EF has a role in determining disease severity in MS, explored by the ARMSS score. As postulated, exposure to such factors specially seems to play a key role when occurring early in the disease history, during adolescence and early adulthood.

7. Study 5: Genetic factors driving chronic silent inflammation

Disability worsening over the years –which is independent of new relapses or new lesions at conventional MRI– is the main clinical feature in patients with progressive MS (Lassmann et al, 2012).

Despite several drugs available for RR-MS, the treatment of the disease in its progressive course is still an unmet need (Feinstein et al, 2015). For this reason, the study of the mechanisms underlying disease progression gathered much attention in the past years.

From a pathological point of view, the chronic lesions represent the most common lesion type in progressive MS (Kuhlmann et al, 2016). In the chronic and active subtype, an ongoing inflammation is observed at the lesion border, despite complete demyelination at the lesion core, and sustains persistent demyelination and axonal injury (Kuhlmann et al, 2016).

In the past few years, the first *in vivo* putative correlates of the chronic active lesions have been discovered. Lesions showing at their borders a paramagnetic rim that is enriched in iron content have been identified by susceptibility imaging techniques at MRI and therefore named as Paramagnetic Rim Lesions (PRL) (Absinta et al, 2016; Absinta et al, 2018). Even though the precise relationship between the PRL and the chronic active lesions is still not completely clear, the PRL are gathering increasing attention from researchers and clinicians, as they are also correlated with disability, progression, and neurodegeneration in MS (Absinta et al, 2019; Maggi et al, 2021). In this perspective, they could represent an ideal candidate outcome in pharmacological clinical trials in the context of progressive MS.

For this reason, a consistent effort was devoted to dissecting the cellular and molecular profile of such lesions. Important findings from single-nuclei RNA sequencing studies led to the identification of a microglia phenotype (the so-called “microglia inflamed in MS”) at the edges of the chronic active lesions, which sustains a silent chronic inflammation, responsible for an outward propagation of the lesion over the years (Schirmer et al, 2019; Jäkel et al, 2019; Absinta et al, 2021). The cells composing the microglia inflamed in MS show consistent differential expression of a variety of genes related to iron metabolism and oxidative stress (Absinta et al, 2021), mirroring from a microscopic point of view the iron enrichment clearly visible at MRI.

Herein, we aim to investigate whether genetic variants in iron metabolism genes impact the risk of disease progression. We discovered that a genetic locus in the Hypoxia-Inducible Factor 1- α (*HIF1A*) gene significantly impacts the risk of developing SP-MS versus RR-MS, affecting the expression of the gene in the immune cells of MS patients, and being associated to an impact in neurofilament light chain (NFL) levels and the burden of PRL.

7.1 Results

7.1.1 Genetic study on iron metabolism prioritizes a locus *HIF1A*

Definition of the study cohort, genetic QC and selection of the SNPs belonging to genes related to iron metabolism (n=23,019) are described in the Methods section (Chapter 9.5). Then, we conducted a genetic analysis in the discovery cohort (n=755) from IRCCS San Raffaele Hospital (OSR), comparing RR-MS versus SP-MS patients (Table 7.1).

We found a statistically significant association, that survived correction for multiple testing, involving SNPs in chromosome 14 and the risk of developing a RR versus SP course (Figure 7.1). Specifically, the minor allele A of the lead variant (rs11621525) was significantly associated with a lower risk of having a SP course, while it was more common in patients with RR-MS ($p=3.30E-06$; $OR_{SP}=0.57$). The minor allele A of the lead variant rs11621525 has an allele frequency in the European population of 0.17 (Karczewski et al, 2020).

This association was successfully replicated in an independent nationwide cohort of 2,062 patients from Sweden (lead SNP p -value= $7.88E-03$, $OR_{SP}=0.79$) as shown in Table 7.2. The lead variant in the Swedish cohort (SWE) is in almost perfect linkage disequilibrium ($R^2=0.98$) with the lead variant of the OSR cohort, representing a proxy of the same signal and supporting the same biological effect.

	<i>OSR</i>	<i>SWE</i>
<i>N</i>	755 (RR: 393; SP: 362)	2062 (RR: 863; SP: 1199)
<i>F/M ratio</i>	2.01	2.83
<i>Age at onset</i>	All: 29.7 RR: 26.51; SP: 32.57	All: 32.58 RR: 28.4; SP: 35.6
<i>Median EDSS at FU</i>	RR: 1.5 (range 0-3.5) SP: 6.5 (range 4-9.5)	RR: 2 (range 0-3.5) SP: 6.5 (range: 4-9.5)

Table 7.1. Clinical and demographic features of the two cohorts involved in genetic association study. F/M = female/male; FU = follow-up.

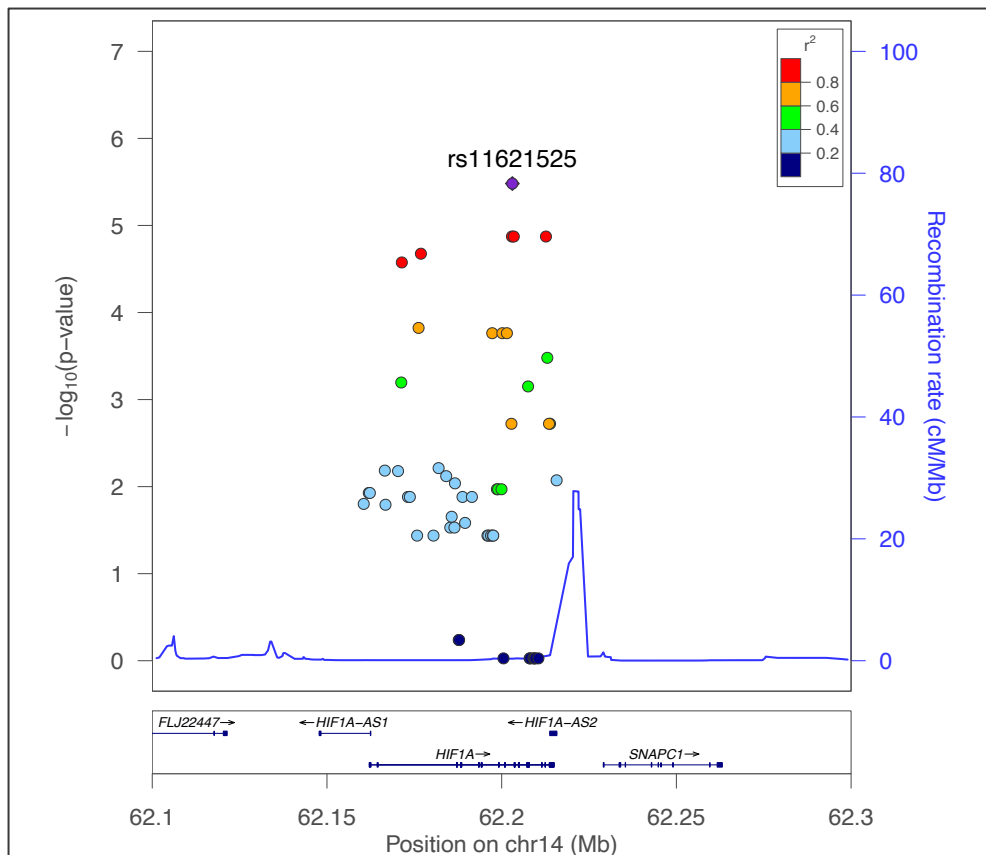


Figure 7.1. Regional plot for the top associated signal in chromosome 14 in the OSR cohort. Colors of the dots reflect the R^2 values, indicating linkage disequilibrium. Reference genome is GRCh37/19.

<i>CHR</i>	<i>SNP</i>	<i>BP</i>	<i>AI</i>	<i>OR.OSR</i>	<i>P.OSR</i>	<i>NMISS.SWE</i>	<i>OR.SWE</i>	<i>P.SWE</i>
14	rs11621525	62203056	A	0.57	3.30E-06	2056	0.802	0.0120
14	rs10873142	62203462	C	0.59	1.34E-05	2042	0.795	0.0080
14	rs4899057	62202942	G	0.59	1.34E-05	2045	0.797	0.0087
14	rs4902082	62212675	C	0.59	1.34E-05	2032	0.798	0.0095
14	rs12435848	62176891	A	0.61	2.11E-05	2053	0.799	0.0087
14	rs1951795	62171426	A	0.60	2.66E-05	2062	0.794	0.0079
14	rs12232182	62176220	C	0.66	1.50E-04	2051	0.866	0.0802
14	rs12434438	62197298	G	0.65	1.73E-04	2061	0.852	0.0546
14	rs2301111	62200201	G	0.65	1.73E-04	2053	0.854	0.0584
14	rs7153817	62201500	G	0.65	1.73E-04	2053	0.854	0.0584

Table 7.2. Top 10 associated SNPs in the OSR cohort and replication in the SWE cohort. For both the cohorts, the OR of having a SP course is reported. NMISS = Number of Non-Missing Genotypes.

The signal maps to chromosome 14 and it is localized in an intron of the gene Hypoxia-Inducible Factor 1- α (*HIF1A*), a fundamental regulator of cell response to hypoxia and iron metabolism, as well as of different immune mechanisms (Zhang et al, 2018).

In literature, multiple evidence pinpoints the functional effects of the variants in this genetic locus, strongly supporting their role on a change of *HIF1A* expression.

Namely, the rs11621525 variant is known to exert an expression Quantitative-Trait-Loci (eQTL) effect on *HIF1A*, as the A allele resulted to be associated to decreased expression of *HIF1A* in whole blood in a large cohort involving more than 31,000 healthy subjects ($p=1.16E-21$) (Vösa U et al, 2021).

A similar effect was found also when assessing the methylation Quantitative-Trait-Loci (mQTL) effect. Notably, results from a meta-analysis from an international consortium involving more than 36,000 healthy subjects (GoDMC, 2021) showed that the rs11621525 A allele is associated to increased methylation levels in whole blood of a CpG (cg20931965) located in the enhancer region for *HIF1A*, therefore strongly suggesting a decrease in gene expression ($p=8.04E-175$). This effect was also present in a smaller study (cg20931965, $p=2.31E-21$) on a Dutch cohort (Bonder et al, 2017), which also found an association between the SNP and increased methylation level of a CpG located in the promoter region of *HIF1A* (cg23174662, $p=6.68E-23$).

7.1.2 eQTL and mQTL in MS patients

To gain an insight on potential similar effects on gene expression and methylation exerted by the variant in the *HIF1A* locus in MS patients, we studied the eQTL effect in PBMC from 78 RR-MS patients who were naive from DMT and had not been treated with steroid drugs in the 30 days before sampling. We found that the rs11621525 A allele was associated with a decreased expression of *HIF1A* ($p=0.0193$, $\beta=-0.54$), adjusting for age and sex (Figure 7.2). This finding confirms that the A allele, which is protective towards the SP course, leads to decreased *HIF1A* expression in the immune cells of patients with MS.

Details regarding PBMC extraction, RNA-sequencing experiments, and eQTL effect analysis are reported in the Methods section (Chapter 9.5.6)

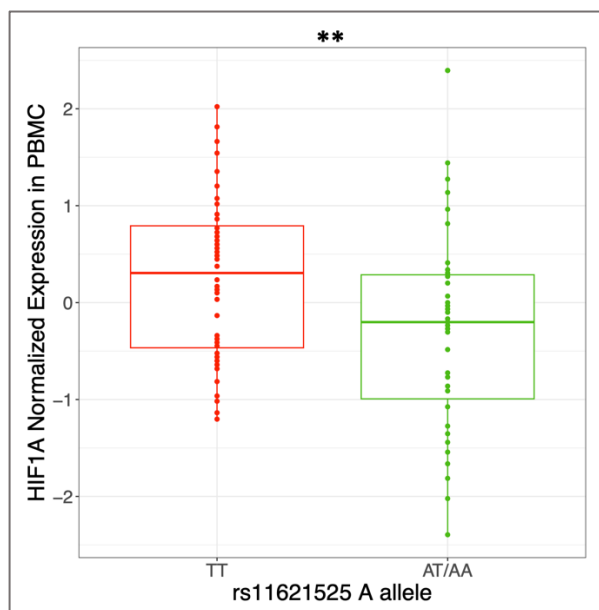


Figure 7.2. Expression Quantitative-Trait-Loci (eQTL) effect in PBMC from naive RR-MS patients. On the x axis the rs11621525 genotype is shown. On the y axis, inverse normal transformed RNA-seq counts for *HIF1A* are shown.

7.1.3 Impact of the *HIF1A* variant on NFL levels

To further assess the relationship between the genetic variation in the reported *HIF1A* locus and silent subclinical inflammation, we also studied the NFL levels, a recognized marker of ongoing axonal injury and chronic white matter inflammation, that have also been found associated with the burden of PRL (Maggi et al, 2021).

Plasma NFL (pNFL) levels were available for a subset of the patients involved in the association analysis of the SWE cohort, that had a RR-MS course at the time of sampling and had been sampled within 20 years from disease onset (n=117). In this group, we found that patients with at least one copy of the rs11621525 A allele showed lower pNFL levels (p=0.0026) (Figure 7.3A).

The same association was found when assessing CSF NFL levels, that were available for 71 RR-MS. The A allele for rs11621525 showed the same effect and was significantly associated with lower NFL (p=0.049) (Figure 7.3B).

A similar effect was found in patients at first demyelinating event (Clinically Isolated Syndrome, CIS), in whom the A allele was associated with lower CSF NFL levels (p=0.019, n=34), further pointing out the impact of reported genetic association on chronic inflammation and ongoing axonal neurodegeneration.

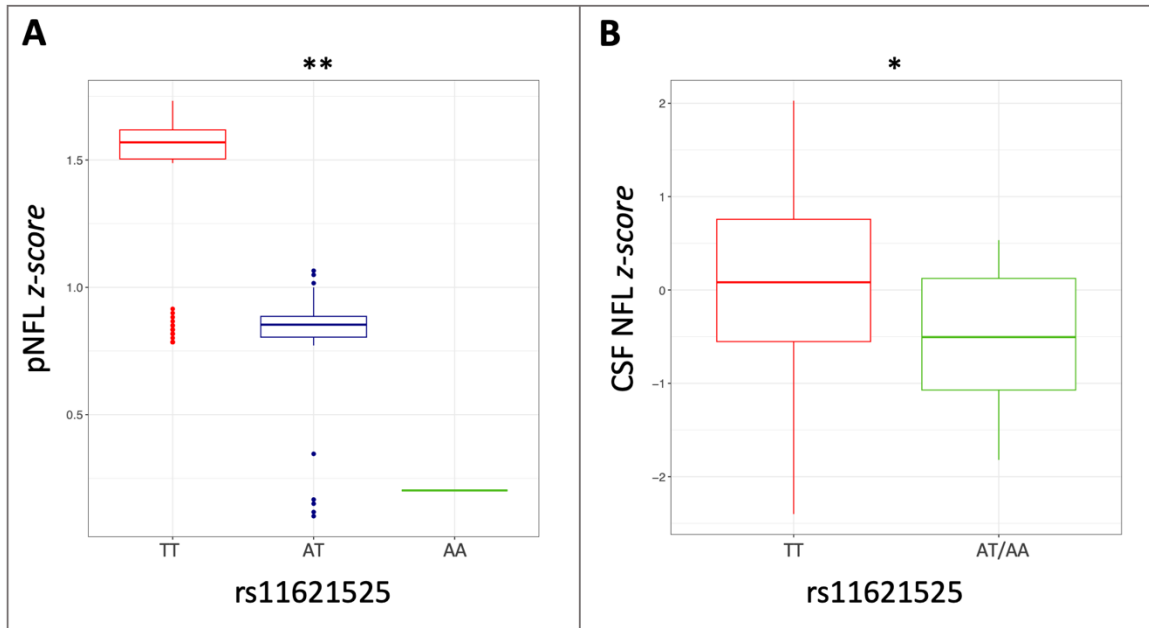


Figure 7.3. Effect of the rs11621525 variant of plasma and CSF NFL levels in RR-MS. NFL levels are reported in form of a z-score, after normalization by rank transformation. * = p -value < 0.05; ** = p -value < 0.01.

7.1.4 HIF1A variant and response to Disease-Modifying Treatment (DMT)

To give an insight on a potential effect of the genetic variant in *HIF1A* on disease-modifying treatment and silent inflammation, we studied whether pNFL levels, which are known to decrease in MS after the beginning of a DMT (De Flon et al, 2016; Kuhle et al, 2020; Bridel et al 2021), were additionally influenced by the rs11621525 A allele.

Overall, in our analysis, the levels of pNFL decreased at 1 year from the beginning of a DMT, as expected. However, we found that in patients starting dimethyl fumarate (DMF), but not other DMT (teriflunomide, fingolimod or natalizumab), the rs11621525 A allele was associated with a more pronounced reduction of pNFL levels after 1 year from treatment start (+/- 6 months), supporting that genetic variation in the *HIF1A* locus may impact the response to DMF treatment ($p=0.027$) (Figure 7.4).

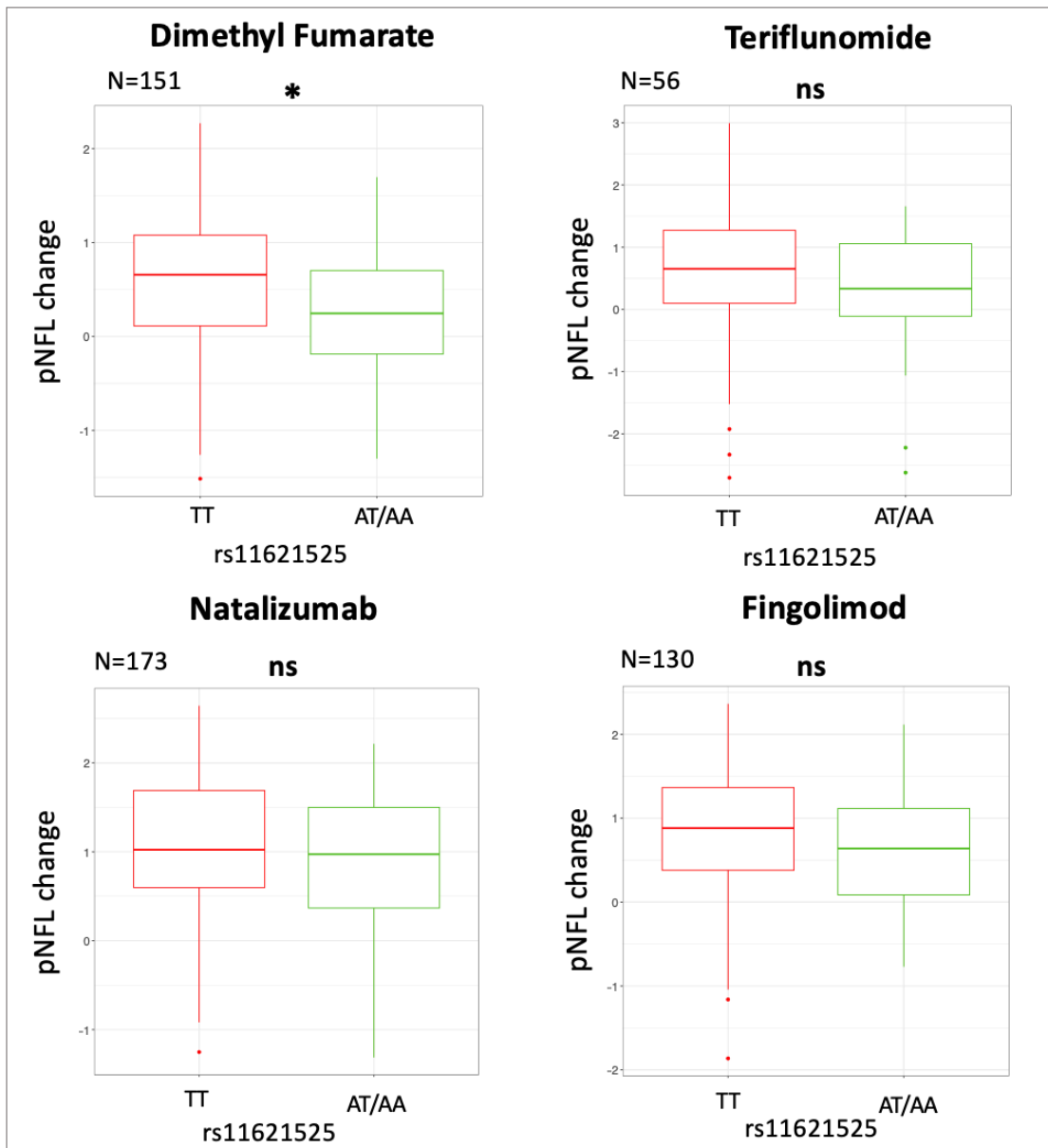


Figure 7.4. Impact of the rs11621525 variant on the NFL levels after DMT. The change in plasma NFL z-score after treatment is shown (y-axis), grouping patients by the status of the rs11621525 A allele (TT versus AT/AA) on the x-axis. * = p-value < 0.05; ns = not statistically significant.

The positive effect of the rs11621525 A allele on the response to treatment with DMF was also confirmed clinically in a cohort of 138 RR-MS patients from OSR on continuous DMF treatment. In this cohort, carriers of the A allele were more likely to reach the NEDA-3 status at 2-years from treatment start, adjusting for sex and age at onset (p=0.0062). This association was even stronger when restricting the analysis to the subset

of patients who were DMT-naive (n=49, p=7.12E-04). Detailed inclusion criteria are reported in Chapter 9.5.9.

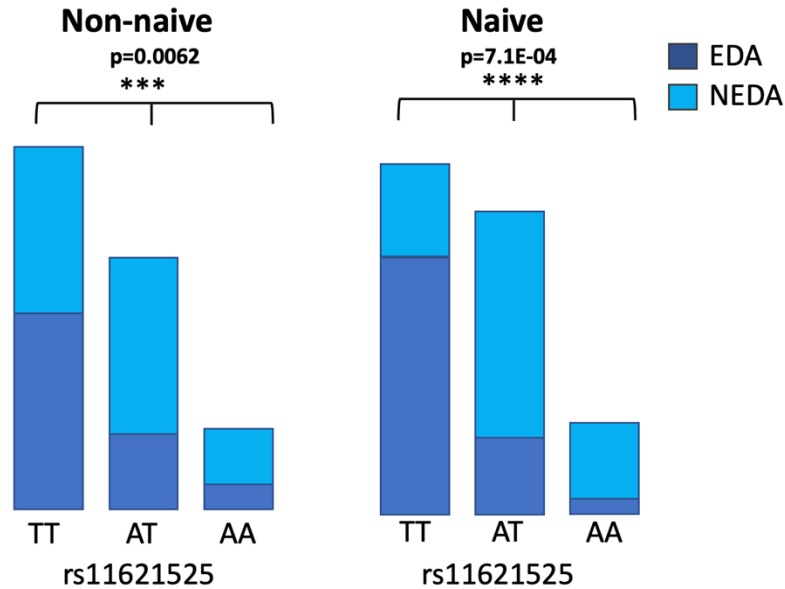


Figure 7.1. Effect of the variant in *HIF1A* on the clinical response to treatment with Dimethyl Fumarate. EDA = Evidence of Disease Activity; NEDA = No Evidence of Disease Activity.

7.1.5 The genetic variant in *HIF1A* affects the PRL burden

As we found a genetic variant in iron metabolism that affects the susceptibility to SP-MS, we also investigated whether the same variant exerts an effect on the PRL, which, as discussed, are characterized by a consistent iron enrichment at their edges where the inflammation becomes chronically active.

We then proceeded with MRI acquisition and PRL segmentation in a subgroup of subjects for which rs11621525 genotype was available, as shown in Figure 7.6 and reported in the Methods section (Chapter 9.5.10).

We found that the A allele for rs11621525 was significantly associated with lower mean volume of the PRL (p=0.0163, beta=-0.60) in a cohort of 35 RR-MS patients, adjusting for sex and total number of PRL (Figure 7.7).

This finding confirms that genetic factors in *HIF1A* are important for the development and expansion of PRL, mirroring an effect of the gene on silent inflammation and disease progression.

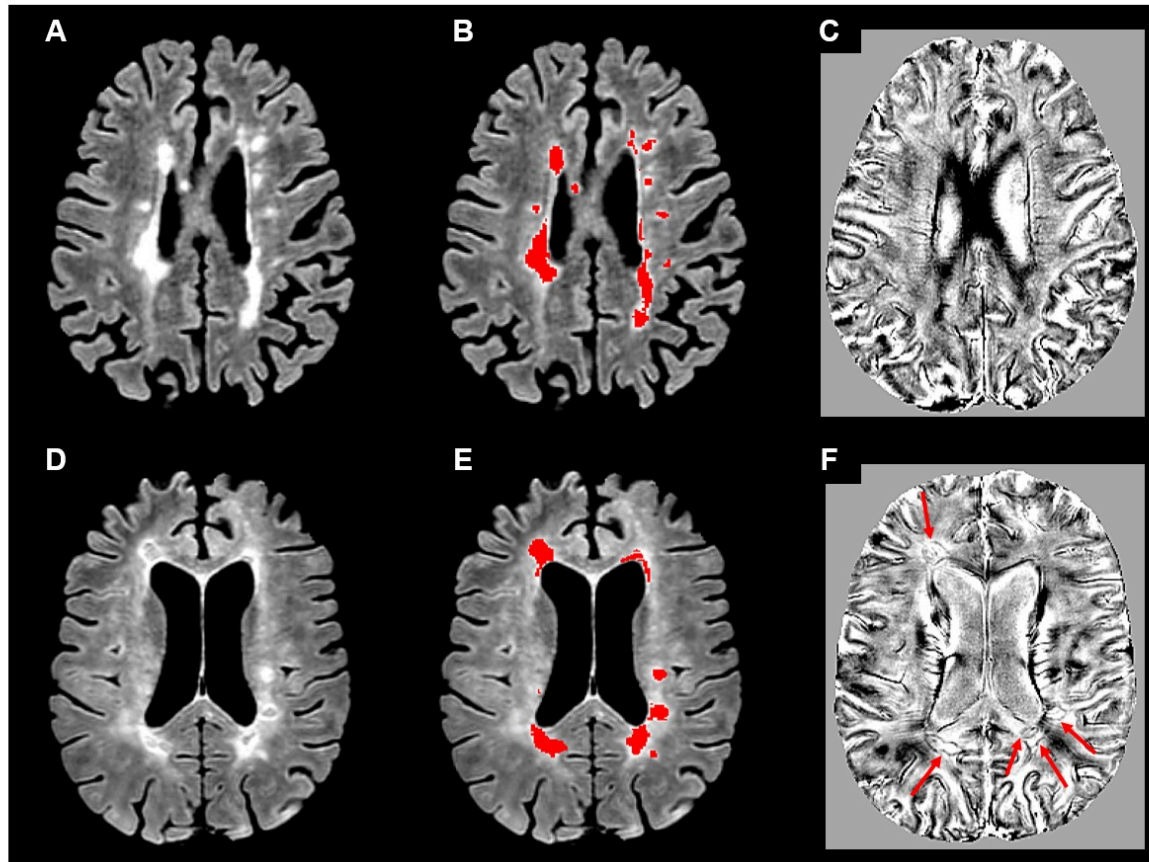


Figure 7.6. Examples of two multiple sclerosis patients with or without paramagnetic rim lesions. A and D: On 3D axial fluid-attenuated inversion recovery sequence, multiple T2-hyperintense white matter lesions are visible in two multiple sclerosis patients. B and C: Brain T2-hyperintense white matter lesions (red-coded) were identified by a fully automated approach on 3D axial fluid-attenuated inversion recovery sequence and their masks were then registered onto the susceptibility-weighted image space. On unwrapped phase images, no paramagnetic rim lesion was found in MS patient #1 (C), whereas several paramagnetic rim lesions (red arrows) were found in MS patients #2 (F).

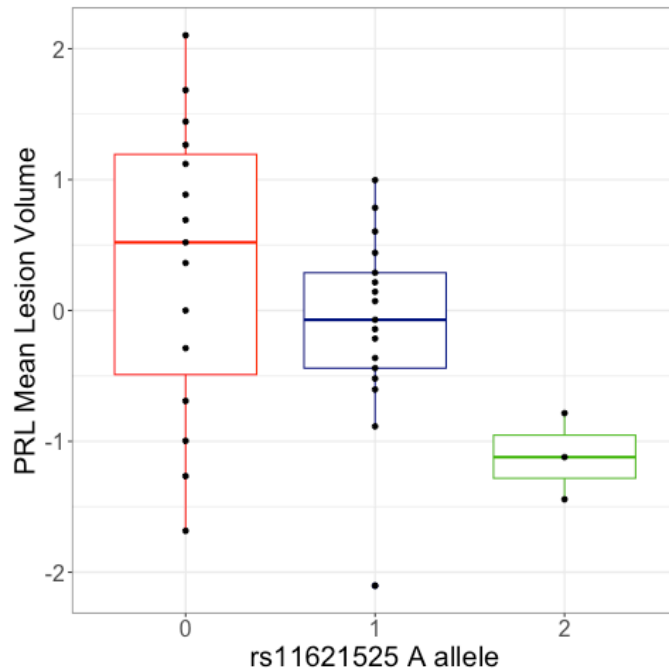


Figure 7.7. The variant in *HIF1A* affects the volumetric burden of the paramagnetic rim lesions. Y-axis: rank-transformed mean volume of the PRL; x-axis = number of copies of rs11621525 A allele.

7.2 Considerations

Iron is an essential cofactor for a variety of enzymes involved in maintaining the health of oligodendrocytes and myelin, and it is involved in remyelination mechanisms (Stephenson et al, 2014). At the same time, iron accumulation is detrimental for cell survival, as it is associated with oxidative stress, development of a pro-inflammatory environment, glutamate toxicity, and impaired DNA repair (Stephenson et al, 2014).

Since iron is a fundamental component of lesions that typically associate with silent inflammation and disease progression in MS, we explored whether genetic factors in iron metabolism could affect the susceptibility to develop progressive MS and could help prioritizing meaningful molecular mechanisms involved in the pathobiology of the PRL and progressive MS.

The results of our genetic analysis implied a locus in the *HIF1A* locus in the susceptibility to SP-MS in a discovery Italian cohort, that was successfully replicated in an independent larger Swedish cohort. Taking advantage of a broad set of data, we then clarified the impact of the lead variant on gene expression, NFL levels, and the burden of

the PRL, adding different levels of evidence that involve HIF1A as a driver of chronic silent inflammation in MS.

The contribution of this gene is also supported by a recent single nuclei RNA-sequencing study, in which *HIF1A* was found to be upregulated in the chronic active edge of MS lesions (Absinta et al, 2021). In the same study, in an analysis comparing different immune cells (excluding lymphocytes), *HIF1A* was upregulated in a cluster of cells that identified as monocytes or monocyte-derived dendritic cells (moDC). Cells from this cluster presented a strong upregulation of all the genes belonging to HIF1A signaling pathway, hypoxia and ferroptosis, and the authors suggested that they could represent a cluster of antigen-presenting cells (APC) which are exposed to a hypoxic environment (Absinta et al, 2021).

The moDC functionally differentiate from monocytes, migrate from the bone marrow to the site of inflammation (Merad et al, 2013) and act as professional APC for memory and helper T lymphocytes, which constitutes the last step in the inflammatory cascade of MS lesions (Hemmer et al, 2015). HIF1A is fundamental for the function and migration of monocytes and moDC. In a low oxygen environment, *HIF1A* is upregulated in the dendritic cells, and impacts their ability to respond to chemokines that drive their migration and activity (Köhler et al, 2012).

HIF1A is essential in the fine tuning of the Th17/T regulatory lymphocytes balance. *HIF1A* knock-out mice do not develop the symptoms of the Experimental Autoimmune Encephalomyelitis (EAE) and show an increase of the T regulatory cells and a decrease of the effector Th17 cells, additionally supporting that the activation of *HIF1A* is key in the context of inflammation (Dang et al, 2011). Importantly, hypoxia triggers *HIF1A* activation not only in T lymphocytes, but also in cells of the myeloid lineage (Cramer et al, 2003).

To gain an insight on a possible effect on the response to MS treatment, we then assessed if the variant in the *HIF1A* locus affects the levels of NFL at one-year after the beginning of a DMT. In this experiment, in patients starting DMF, but not other DMT, the rs11621525 A allele significantly improved the response to the drug, in terms of NFL reduction. We also explored this finding on a clinical level in a pharmacogenomic analysis that additionally supported that the variant significantly affects the response to DMF treatment.

DMF, an exogenous fumaric acid ester, is an approved drug for the treatment of RR-MS and has multiple effects, involving both peripheral immunity and the central nervous system (Blair H, 2019). In the periphery, DMF reduces the activation of T helper (Th) lymphocytes in the effector Th17 cells and enhances the T regulatory and Th2 counterpart (Wu et al, 2017). In the central nervous system, DMF exerts an antioxidant activity, acting through the Nuclear Factor Erythroid 2-Related Factor 2 (Nrf2) and the upregulation of genes as SOD2, ferritin and GST (Mills et al, 2018). Moreover, monomethyl fumarate, the active metabolite of DMF, crosses the blood-brain barrier and reduces glutamate toxicity (Luchtman et al, 2016). Therefore, a neuroprotective of DMF was hypothesized, even though the exact mechanisms have not been fully explored and understood so far.

The pharmacodynamics of DMF has many points in common with mechanisms involving HIF1A. Endogenous fumarate is one of the main regulators of *HIF1A* expression (Isaacs et al, 2005) and DMF increases the degradation of HIF1A in the proteasome (Zhao et al, 2014). The inhibition of *HIF1A* leads to increased NRF2 expression, promoting an antioxidant activity (Briggs et al, 2016).

HIF1A is overexpressed by the moDC and in the edges of the chronic active lesions (Absinta et al, 2021), and monocytes have been found to be significantly affected by DMF treatment, together with production of Reactive Oxygen Species (ROS) (Carlström et al, 2019), providing additional evidence that myeloid lineage cells are crucial in the pathogenesis of MS (Galli et al, 2019). In this regard, preliminary data show that anti-CD20 treatment, that conversely acts on cells of the lymphoid lineage, does not affect PRL changes, but further validation is strongly needed (Maggi et al, 2022).

Exploring the mechanisms driving disease progression in MS is crucial, as the treatment of the disease in its progressive course is still largely an unmet need. The PRL are the most promising *in vivo* biological correlate of silent inflammation in MS and an increasing number of clinical trials is introducing them as an outcome measure, when testing the efficacy of drugs on disease progression (Martire et al, 2022).

Our data show that genetic variation in the *HIF1A* gene impacts silent inflammation in progressive MS patients. This novel mechanism is strictly linked to endogenous and exogenous fumaric acid esters, prompting future investigation to assess their potential in the treatment of chronic inflammation in MS.

8. Discussion

In the present work, we took advantage of multiple layers of information to explore disease heterogeneity in MS, through a diverse set of measures of disease activity, severity, and progression. The results of the single studies are discussed more extensively in the respective chapters. Herein, we want to provide some general considerations that emerge from our work, when integrating different kinds of evidence from different studies.

In a genome-wide meta-analysis involving 1,408 patients with MS, we found different genetic loci that showed a potential association with disease activity, assessed by the NEDA-3 status at 2-years from the beginning of a first line DMT. Some of the implicated genes take on a special interest in the context of MS mechanisms, as the *SEMA3E* and the *NAMPT* loci, which are discussed in Chapter 3.4.

In the same study, the second top associated locus mapped to the *MKRN3* (Makorin Ring Finger Protein 3) gene. This gene is well known as a regulator of the Gonadotropin-Releasing hormone and acts as a brake for the onset of puberty, modulating hypothalamic activity (Abreu et al, 2020). At the end of childhood, as puberty approaches, the expression of *MKRN3* starts to decrease in brain, eventually resulting in increased gonadotropin expression (Abreu et al, 2020). During puberty, there is a remodulation of the levels of the sex hormones, accompanied by the onset of reproductive maturity and the development of the secondary sexual characteristics (Wood et al, 2019). Puberty is also known to lead to extensive changes in the immune system, affecting the T regulatory balance, the maturation of dendritic cells and the formation of IgG antibodies (Ucciferri and Dunn, 2022). Since MS is more likely to first present in young adults in the reproductive age, different works have explored the role of puberty and changes in circulating levels of sex hormones in MS (Chitnis, 2013). Multiple levels of evidence show that factors regulated by sex hormones, that dramatically change after the onset of puberty, are important in the disease, in which the epigenetic component has a primary role, even though the exact mechanisms are not clear (Thompson EE et al, 2018; Ucciferri and Dunn, 2022). In this work, we found an association between a genetic locus in *MKRN3*, that is known to regulate the onset of puberty, and disease activity in MS.

Remarkably, when investigating how the methylation profile in the immune cells impacts the same outcome of disease activity, the NEDA-3 status, again our analysis

supported the importance of another fundamental regulator of puberty and sexual maturation, that is the AMH. The role of sexual differentiation and maturation was also additionally fostered by a Gene Ontology enrichment analysis, that prioritized terms as ‘*Fertilization*’ and ‘*Regulation of reproductive system*’ (Table 4.6). Moreover, in an analysis on 1,117 women with MS we found that the use of oral contraceptive, menopause and pregnancy before the diagnosis were all associated with a less severe disease course of MS, together with a trend for association between an older age at menarche and less severe MS (Table 6.2). We reckon that the evidence coming from this integrated approach of different ‘*omics*’ (genomics, epigenomics and environmental factors) raises interest towards the role on puberty on disease manifestations, in terms of disease activity and global severity. In this regard, our work confirms that the extensive modifications that occur in this age are a crucial step in the cascade of events that lead to MS.

In Chapter 7, we found evidence that a locus in the *HIF1A* gene affects the susceptibility to progressive MS in two independent cohorts from Italy and Sweden, for a total of 2,817 patients. In the same Chapter, through additional multiple levels of evidence, we depicted how HIF1A, likely finely modulated by genetic factors, acts as a strong driver of the chronic and silent inflammation that is characteristic of the progressive course in MS. We were also able to correlate genetic variation in *HIF1A* with the paramagnetic rim lesions, the biological *in vivo* correlate of chronic inflammation. Surprisingly, in an independent double-centric cohort of patients described in Chapter 4, we found that increased methylation levels in a CpG in the body of the *APBA3* (Amyloid Beta Precursor Protein Binding Family A member 3) gene was significantly associated to higher probability of having signs of disease activity, and this association held after correction for multiple testing (Table 4.3). *APBA3* is a potent inducer of *HIF1A* in monocytes and macrophages in hypoxic conditions, promoting their inflammatory activity and their migration to sites of tissue inflammation (Hara et al, 2017). We reckon that it is remarkable that, in two independent studies investigating manifestations of disease under different circumstances, *HIF1A* emerged as a hub that drives inflammation in MS at different stages. We hypothesize that different measures can capture a single biological entity, which is probably active since the early phases of MS, as also supported by recent literature (Kaufmann et al, 2022; Portaccio et al, 2022).

Of note, HIF1A is also the connecting element between inflammation and oxidative stress. We discovered that the response to Dimethyl Fumarate, a DMT for MS which is known to play a key role in the modulation of oxidative stress (Carlström et al, 2019), is affected by genetic variation in *HIF1A*. This preliminary evidence fosters future investigation on the potential of Dimethyl Fumarate in hampering chronic inflammation in progressive MS.

Notably, in Chapter 5 we found a protective effect of alcohol on disease severity, assessed by the ARMSS score. In a sub-analysis the protective effect was found if patients reported to drink wine or beer, while the effect of spirits consumption was detrimental (Table 6.4). Even though a validation in larger cohorts is needed to rule out any potential confounder, this finding may additionally support that antioxidant activity on the central nervous system, which is exerted by different compounds as the polyphenols contained in wine, may have a benefit on disease severity (Zhou et al, 2021). In this regard, it is also worth mentioning that, in Chapter 5, we provided evidence that higher vitamin D levels are causally associated with disease activity in MS. Vitamin D, along with many functions that primary act on the immune system, was hypothesized to play a role also on the central nervous system through neuroprotective and antioxidant mechanisms implicating Nrf2 (Nachliely et al, 2019), a transcription factor which is strictly linked to *HIF1A* and the Dimethyl Fumarate, as described in Chapter 7.

To conclude, in this PhD project we found different molecular and environmental markers associated with the diverse manifestations of the disease, exploring the spectrum of clinical heterogeneity in MS at multiple levels. We showed that an integration of different sources of evidence, coming from molecular, clinical, and environmental data, can be successful in identifying mechanisms that are relevant to disease pathogenesis.

We strongly believe in the value of this integrated approach. Our focus for future investigation will be translating our discoveries to meaningful lifestyle and pharmacological interventions.

9. Methods

9.1 Study 1: Genetic factors underlying disease activity

9.1.1 Inclusion criteria and NEDA-3 definition

The Italian cohort was composed by patients who had been admitted to the Multiple Sclerosis Centre of the IRCCS San Raffaele Hospital in Milan (OSR). We conducted a manual revision of medical records and databases to collect the data which are necessary to define the NEDA-3 status after two-years follow-up. For each patient, as a baseline (BL) timepoint, we selected the beginning of the earliest first line DMT for which complete clinical data were available to define the NEDA-3 status after two years from treatment start. We excluded patients who had been treated with immunosuppressive (e.g., azathioprine, mitoxantrone, cyclophosphamide, etc.) or second line (e.g., ocrelizumab, natalizumab, fingolimod, cladribine, alemtuzumab, etc.) drugs prior to BL timepoint.

Patients were classified as reaching the NEDA-3 status after 2 years from the beginning of a first line DMT (interferon, glatiramer acetate, dimethyl fumarate or teriflunomide), if all of the following were satisfied during the follow-up time: 1) no relapses; 2) no new or enlarging T2 lesions and no gadolinium-enhancing at brain or spinal cord MRI; 3) no Expanded Disability Status Scale (EDSS) score progression. EDSS progression was defined as follows: an increase in the EDSS score of at least 1.5 if baseline EDSS was 0; an increase of EDSS of at least 1.0 if baseline EDSS was between 1 and 5.5 points; an increase of EDSS of at least 0.5 points if baseline EDSS was ≥ 6.0 .

In order to be included, patients also needed to have available whole-genome genetic data at the Laboratory of Human Genetics of Neurological Disorders in OSR.

For the OSR cohort, a total of 1,183 patients with available clinical data to define NEDA-3 status at 2-year follow-up and available genetic data were included in the study and underwent the quality controls (QC).

In the French cohort, patients were recruited at the Centre Hospitalier Universitaire de Toulouse (CHUT), according to the same inclusion criteria used for the OSR cohort. Harmonization of the clinical datasets involving the two cohorts was performed at OSR. A total of 299 patients with available clinical and genetic data underwent the QC.

9.1.2 Genotype imputation and QC in the OSR cohort

Imputation to reference genome (Haplotype Reference Consortium, McCarthy et al, 2016) was conducted on all the individuals with available whole-genome genetic data at the Laboratory of Human Genetics of Neurological Disorders in OSR.

Since individuals had been genotyped on four different platforms (Illumina OmniExpress, Illumina Omni 2.5, Illumina Human Quad, Illumina Global Screening Array), imputation was carried out separately on each platform with later merging on bona-fide imputed variants. Given the high level of overlap between OmniExpress and Omni2.5, samples genotyped on these two platforms were jointly quality-controlled and imputed. Variants' rsIDs were assigned from dbSNP v151 GRCh37p13 (<http://www.ncbi.nlm.nih.gov/SNP/>).

Prior to imputation, a set of homogeneous QC steps at sample and SNP level were conducted for each of the three cohorts separately. At sample level, we excluded subjects for which a mismatch with declared sex was observed, those with call-rate < 90% and outliers exceeding the mean level of heterozygosity by >3 standard deviations. At variant level, we discarded rare SNPs with AF<1%, SNPs with a call-rate<90% and those departing from HWE at $p < 10E-6$.

After imputation, we extracted the subjects who had fulfilled the clinical inclusion criteria, for a total of 1,170 subjects and 9,388,047 variants that passed imputation QC. On these, we applied an additional round of study-specific QC. To identify individuals with potential biological relationship in the study cohort, we performed a pairwise identical by descent (IBD) estimate, that found 5 pairs of subjects with potential relatedness ($PI_HAT \geq 0.250$). Two pairs were excluded for $PI_HAT \geq 0.50$ and of the remaining 3 pairs, the individuals who were the least characterized were removed (total number of individuals removed: 7). To identify population outliers, we run a Principal Component Analysis (PCA) and excluded nine subjects (mean +/- 6 standard deviations) (Figure 3.1). As an additional level of QC for the study cohort, we retained only the SNPs with a call rate $\Rightarrow 0.99$ and Minor Allele Frequency (MAF) of at least 5%. A PCA on the final dataset was also run to use PCs as covariates in the subsequent association analysis. The final dataset included 1,154 subjects and 4,231,855 variants.

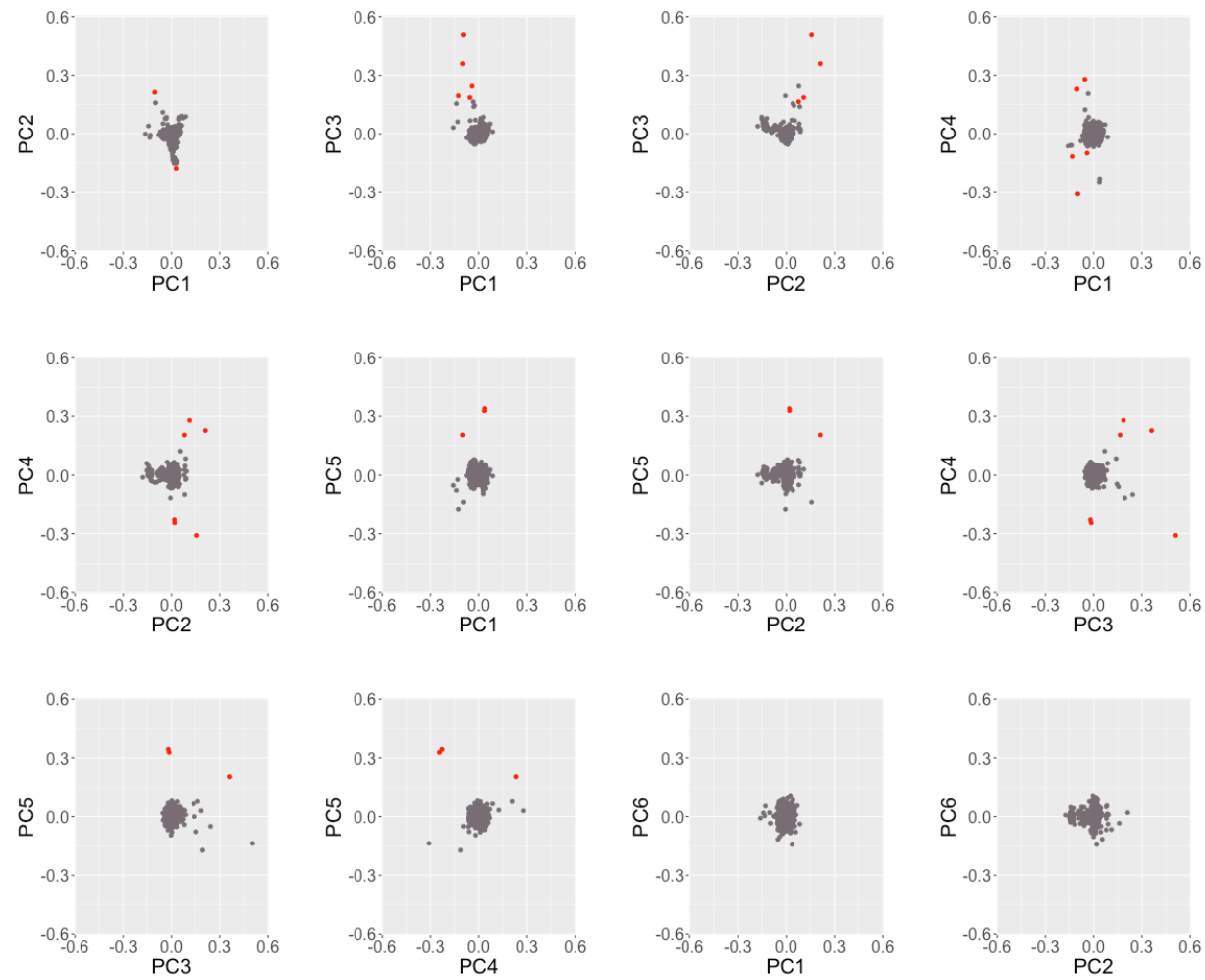


Figure 9.1. Principal Component Analysis for the OSR cohort. The red points show population outliers (± 6 standard deviations). PC = Principal Component

9.1.3 Genotype imputation and QC in the CHUT cohort

Patients of the CHUT cohort had been genotyped on two different platforms: Axiom Genome-Wide LAT 1 (Affymetrix, USA) and Illumina OmniExpress. The imputation was carried out separately, as described in the QC section for the OSR cohort. Post-imputation QC were performed as reported for the OSR cohort. Starting from 299 patients, a total of 254 subjects and 5,127,516 SNPs were retained after the QC.

9.1.4 Genome-Wide Association Study (GWAS) and meta-analysis

The two cohorts were first analyzed separately using PLINK v1.9 (Chang et al, 2015). Logistic regression was carried out using the NEDA-3 status as outcome and genotype as predictor. To select the covariates for logistic regression model we conducted a preliminary exploratory analysis between clinical and demographic factors and the NEDA-3 status (Table 3.2). Sex, age, and Principal Component (PC) 1 to 8 were included in the OSR cohort as covariate. In the CHUT cohort PC1 and PC2, age and sex were included as covariates. The genetic association analysis was conducted calculating the odds ratio (OR) of having Evidence of Disease Activity (EDA) given by the A1 allele (effect allele).

	<i>OR_{EDA} [95% CI]</i>	<i>p-value</i>
<i>Sex (F)</i>	1.40 [1.09-1.82]	0.0092
<i>Age at onset</i>	0.97 [0.96-0.98]	< 0.001
<i>Age at BL</i>	0.97 [0.96-0.98]	< 0.001
<i>Disease duration</i>	0.99 [0.97-1.01]	0.383
<i>EDSS at BL</i>	1.15 [1.00-1.31]	0.048
<i>Relapses in the 2-yr before BL</i>	1.14 [0.99-1.29]	0.054
<i>DMT before BL</i>	1.21 [0.74-2.04]	0.46

Table 9.2. Covariate analysis of clinical and demographic factors towards the NEDA-3 status. Mann-Whitney or chi-square test is reported on the right column. *OR_{EDA}* = Odds Ratio of having Evidence of Disease Activity (EDA). *BL* = baseline. *EDSS* = Expanded Disability Status Scale score. *DMT* = Disease-Modifying Treatment. 2-yr = 2 years.

After the two single cohort GWAS, the results were meta-analyzed using a fixed-effect model with PLINK v1.9, retaining only the variants that were common to both the cohorts (n=3,948,158). Threshold for genome-wide significance was set at $p < 5e-08$.

9.1.5 Variant to genes

To assign a gene to each of the investigated variants, we used the Variant2Gene pipeline from Open Target Genetics (Ghoussaini et al, 2021; Mountjoy et al 2021). For each variant, this approach allows to calculate a score, representing the probability that a variant exerts its effect on a specific gene. As a novelty, the score integrates not only positional information (distance from a Transcription Start Site), but also extensive evidence from literature involving multiple Quantitative-Trait-Loci (QTL) effects (namely: expression-, proteomic- and splicing- QTL), link with promoter capture Hi-C in 17 human primary hematopoietic cell types, known variant-trait association etc.

9.1.6 Gene ontology enrichment and pathway analysis

Starting from the results of the two-cohorts meta-analysis, we selected the top 100 mapped genes according to Variant2Gene as mentioned, and we used WebGestalt (Liao et al, 2019) to run an enrichment analysis based on Gene Ontology terms in the ‘*Biological process*’ category and a pathway analysis using KEGG. The settings for both analyses were: a range of genes for term/category comprised between 5 and 2,000; a Benjamini-Hochberg correction for multiple testing.

9.2 Study 2: Epigenetic factors underlying disease activity

9.2.1 Study cohort

In this study, we included patients with diagnosis of RR-MS from two different centers: IRCCS San Raffaele Hospital in Milan, Italy (OSR) and Centre Hospitalier Universitaire de Toulouse, France (CHUT). All patients were not on any Disease-Modifying Treatment (DMT) at the time of sampling. Patients previously treated with second line DMT or immunosuppressive drugs were excluded, as well as patients who had been treated with corticosteroid drugs in the 30 days before sampling. All patients started a treatment with a first line DMT (Table 4.1) and were followed up to assess the NEDA-3 status at 2-years from drug start.

A total of 249 patients fulfilling the above-mentioned criteria were included and underwent down-stream methylation profiling and analysis (Table 4.1).

9.2.2. PBMC extraction and methylation profiling

After blood sampling, Peripheral Blood Mononuclear Cells (PBMC) were extracted through density gradient centrifugation by Lymphoprep™ (STEMCELL Technologies, Vancouver, Canada), as described in the manufacturer's protocol. Bisulphite conversion and methylation profiling was performed at the Laboratory of Human Genetics of Neurological Disorder at OSR. A randomization procedure for phenotype (NEDA-3) and center (CHUT/OSR) was carried out to minimize potential batch effect. After bisulphite conversion, whole-epigenome methylation profile was obtained by the means of Illumina Infinium MethylationEPIC BeadChips (Illumina, Inc., San Diego, CA, USA), following the manufacturer's instructions.

9.2.3 Quality Controls (QC)

A total of 249 subjects and 865,859 probes underwent QC. The following probes were filtered out: probes with a detection p-value > 0.01, non-CpG probes, X and Y chromosome probes, cross-reactive probes, probes mapping to known SNPs, probes with less than 3 bead counts in at least 5% of the subjects (Pidsley et al, 2016). No issues were detected during QC on samples when checking for sex or age mismatch and overall sample quality.

After QCs, a total of 249 subjects and 778,879 probes were retained. Within-array normalization and normalization for type I and II probes was performed using the *ssNoob* method, implemented in the *minfi* pipeline (Aryee et al, 2014). M-values ($\log_2 \frac{\beta}{1-\beta}$) were calculated, to be used in statistical analysis (Du et al, 2010). We run a Principal Component Analysis (PCA) on methylation and detected a batch effect exerted by methylation slide and position in the slide, as extensively reported by others (Sun et al, 2011; Ross et al, 2022). Therefore, we applied a batch correction using an empirical Bayes method by the means of the *ComBat* function implemented in the *SVA* package in R (Leek et al, 2022), which reduces error estimates and improves reproducibility (Leek et al, 2010).

In addition, we obtained estimated cell proportions based on methylation beta-values using a reference-free deconvolution approach (Houseman et al, 2016), implemented in the *EpiDISH* R package (Zheng et al, 2018). Cell-proportions (CD8, CD4, monocytes, NK, B cells and neutrophils) were used as a covariate in the subsequent analysis, to minimize cell fraction-dependent differential methylation.

A schematic overview on the QC carried out at the individual- and probe-level is reported in Table 9.2.1.

9.2.4 Differential methylation analysis

We used the *limma* Bioconductor package version 3.50.3 (Ritchie et al, 2015) to build generalized linear models assessing the relationship between DNA methylation and the NEDA-3 status, using as covariates sex, age at disease onset, center, and cell type proportions. The p-values of association were then adjusted by a 5% False-Discovery Rate (FDR) correction to minimize multiple testing bias for the Differentially Methylated Positions (DMP) analysis.

We then built Differentially Methylated Regions (DMR) using the *DMRcate* pipeline (Peters et al, 2019). We used standard parameters ($\lambda = 1000$, $C = 2$), to detect DMR encompassing minimum 3 CpGs, with at least a 2% change in methylation beta-values. DMR were defined by minimum smoothed FDR p-value $< 1e-07$. The identified DMR were ranked using Fisher's combined probability test, as described in the *DMRcate* pipeline (Peters et al, 2019).

9.2.5 Gene ontology enrichment analysis

Enrichment for Gene Ontology (GO) terms (category '*Biological process*') was performed using the top 100 mapped genes from the DMP analysis. The analysis was conducted using WebGestalt (Liao et al, 2019), with standard settings.

9.2.6 MQTL and eQTL effect

Methylation- and expression- Quantitative Trait Loci (mQTL and eQTL) effects were calculated on PBMC from MS patients. RNA-sequencing experiments and related QC are described in Chapter 9.5.6. MQTL and eQTL effects were computed using linear models and adjusting for center, age at sampling, and sex.

<i>Pipeline</i>	<i>Subjects removed</i>	<i>Subjects retained</i>	<i>Probes removed</i>	<i>Probes retained</i>
<i>Overall sample quality</i>	0	249	-	865,859
<i>Sex and age mismatch</i>	0	249	-	865,859
<i>Detection p-value > 0.01 in > 5%</i>	-	-	3,001	862,858
<i>Non-CpG probes</i>	-	-	2,931	859,927
<i>Probes in X or Y chromosome</i>	-	-	19,091	840,836
<i>Cross-reactive or SNP probes</i>	-	-	59,355	781,481
<i>Probes with 3 < BC in => 5% of samples</i>	-	-	2,602	778,879
<i>Final number</i>		249		778,879

Table 9.2.1 Overview of the QC at the individual- and probe-level on the methylation profile. BC = bead counts.

9.3 Study 3: Vitamin D and disease activity

9.3.1 Study cohort

Vitamin D levels were measured in a subgroup of the cohort of untreated MS patients described in Chapter 4 and including patients from IRCCS San Raffaele Hospital (OSR) and the Centre Hospitalier Universitaire de Toulouse (CHUT). As already reported in Chapter 4, the inclusion criteria were a) diagnosis of RR-MS; b) no DMT at the time of blood sampling; c) no previous treatment with second line or immunosuppressive drugs; c) no steroid treatment in the 30 days prior to blood sampling; d) availability of clinical data to evaluate the NEDA status at 2 years from the beginning of a new first line DMT.

Vitamin D measurement was available in a total of 230 patients, fulfilling the above-mentioned inclusion criteria. Clinical features of the cohort are reported in Table 5.1.

9.3.2 Vitamin D measurement and analysis

Vitamin D levels were examined through the measurement of 25-hydroxyvitamin D at CHUT using a commercially available electro-chemiluminescence competitive binding assay (Cobas, Roche) on a Roche Cobas 8000 analyzer. Vitamin D measurement was performed in 230 patients with available information on NEDA-3 status after 2-years of follow-up. In the quality controls, an extreme outlier was excluded from further analysis.

To control for seasonal variation of the vitamin D levels, we modelled a periodic function to estimate the seasonal effect, as suggested (Saltytė Benth et al, 2012):

$$-\sin\left(2\pi * \frac{Month}{12}\right) - \cos\left(2\pi * \frac{Month}{12}\right)$$

Then, we built a linear regression model (Unadjusted vitamin D ~ seasonal effect) and summed the residuals of this model to the unadjusted vitamin D levels, to obtain season-adjusted vitamin D levels, which was used for downstream analysis (Figure 9.3.1). This approach has been previously shown to be effective in minimizing the effect of the seasonal variation and it has already been implemented in other studies on vitamin D in MS (Saltytė Benth et al, 2012; Martinelli et al, 2014).

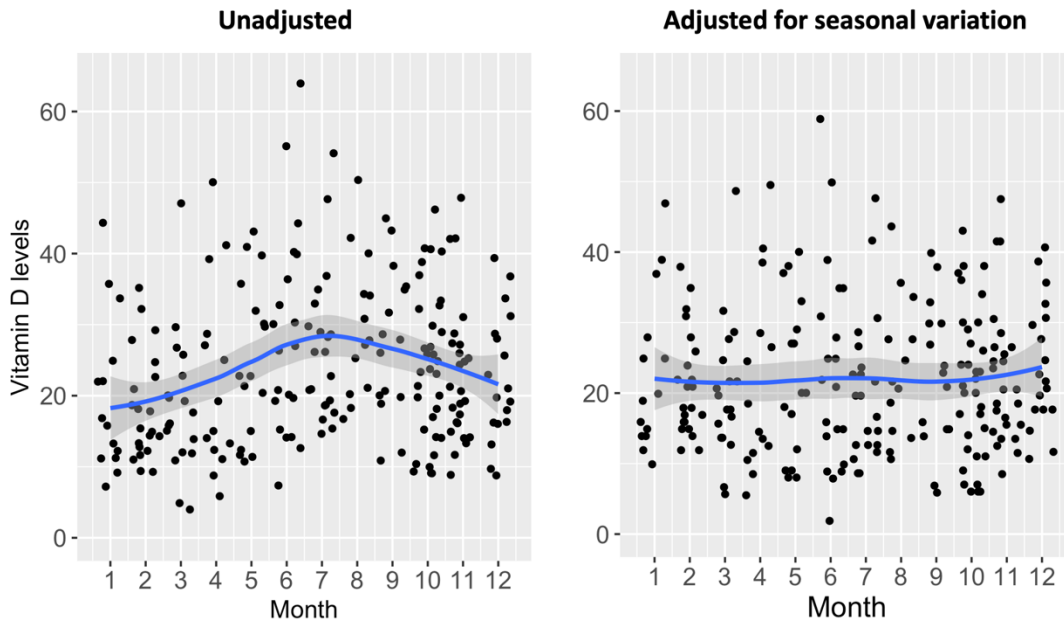


Figure 9.3.1. Effect of the adjustment for the seasonal variation on the vitamin D levels. Vitamin D levels on the y axis are reported in ng/ml. Month: calendar month of blood sampling for vitamin D. A The blue line shows loess smoothing, capturing the seasonal variation in vitamin D levels, which are higher during spring and summer.

A logistic regression model was used to assess the predictive value of baseline adjusted vitamin D level towards the NEDA-3 status after two years from follow-up in non-treated MS patients, adjusting for age at sampling. A Mann-Whitney-Wilcoxon test was used to assess the difference of the mean adjusted vitamin D levels grouping patients by number of relapses during the follow-up (0 = no relapses; 1 = one relapses; 2+ = two or more relapses). Introducing gender or the number of relapses before baseline did not have a significant impact on the results in a subsequent sensitivity analysis.

9.3.3 Mendelian Randomization analysis

For the Mendelian Randomization (MR) analysis, we used the publicly available summary statistics from a previously published GWAS on vitamin D levels (Revez et al, 2020). In this study, vitamin D levels were analyzed in 417,580 subjects of European ancestry from the UK Biobank (Revez et al, 2020) in the form of rank-based inverse normal transformed 25-OH-vitamin D without adjustment for Body Mass Index (BMI). The same analysis was also conducted with vitamin D levels in a form of a 1-standard deviation variation in natural-log transformed values on a slightly lower number of

subjects (416,247) but with BMI adjustment. For our purpose, we used the results of this second analysis with BMI adjustment to exclude its potential confounding effect.

Prior to the MR analysis, we selected the instrumental variables (IV) from the meta-analysis GWAS on NEDA-3 described in Chapter 3. First, we aligned the effect allele in our meta-analysis to match the effect allele reported in the vitamin D GWAS from Revez et al. Then, we retained from our results the variants which satisfied the following criteria:

- same direction of effect in the vitamin D GWAS and in our meta-analysis (that is, increase in vitamin D and higher probability of NEDA3- status at 2-years or decreased in vitamin D and lower probability of NEDA-3 status);
- a genome-wide significant ($p < 5e-08$) association in the vitamin D GWAS and a nominally significant ($p < 0.05$) association in our meta-analysis.

A total of 5,508 variants, before linkage disequilibrium (LD) clumping, satisfied such criteria and underwent downstream analysis. To perform MR analysis, we used the *TwoSampleMR* R package (Emani et al, 2018). As a first step, implemented in the *TwoSampleMR* pipeline, we applied a LD clumping on the 5,508 IVs, as it is a necessary condition for MR to prevent bias. To exclude horizontal pleiotropy (=presence of pathways other than the exposure that explain the effect of the genetic variants on the outcome), we performed a MR Egger analysis and visually inspected the related funnel plot (Figure 5.4D), satisfying the requirement for MR analysis. The final number of IVs selected was 15. To assess the effect of the IVs on the outcome, a random-effects inverse-variance weighted (IVW) analysis was performed. We also applied the following additional methods for MR to validate our primary analysis with the IVW MR: MR Egger, weighted median, weighted mode, simple and weighted mode with the No Measurement Error in the SNP effect (NOME) assumption (Bowden et al, 2015). A leave-one-out approach was used to explore the impact of the exclusion of the most associated IVs on the results.

9.4 Study 4: Environmental factors and disease severity

9.4.1 Study cohort

A total of 1,892 consecutive patients who were admitted to the Multiple Sclerosis Centre at IRCCS San Raffaele Hospital between September 2019 and November 2021 filled out an Environmental Questionnaire (EQ). Detailed clinical data were available for 1,777 out of 1,892 patients. We excluded patients with a follow-up time from onset < 2 years and patients with age < 18 years old at the latest follow-up, for a total of 1,688 patients who were included in the study. The clinical features of the cohort are reported in Table 6.1.

9.4.2 Environmental questionnaire

The EQ was available both in a paper and electronic version, that patients filled out at the Multiple Sclerosis Centre at IRCCS San Raffaele Hospital during follow-up visits or remotely from home, without any supervision. The electronic version was created using LimeSurvey (Limesurvey GmbH). The questionnaire was administered in the local language (Italian) and it was divided into seven sections: a) demographic data, b) personal and family medical history, c) smoking habits, d) sunlight exposure, e) body weight and dietary habits, f) pregnancy and hormonal factors and g) infectious agents and vaccines. For the purpose of this study, we will focus on sections E and F. An extract of EQ as an example is shown in Appendix: Supplementary Figures, Figure S6.1.

9.4.3 Statistical analysis

The analysis was conducted using R version 4.1.3. The latest available Expanded Disability Status Scale (EDSS) score and the age at the time of EDSS evaluation were used to calculate the Age-Related Multiple Sclerosis Severity (ARMSS) score, a validated tool to assess disability in MS patients (Manouchehrinia et al, 2017). The global ARMSS score was calculated on the final cohort (n=1,688) using the package *ms.sev* version 1.0.4 (Westerlind et al, 2016). The ARMSS scores underwent rank transformation to fit a normal distribution (Shapiro-Wilk test p-value before normalization < 0.0001) and were used as an outcome in linear regression models. To assess the potential impact of other factors on the ARMSS score, we also performed a covariate analysis. In particular, neither gender (p=0.40) or SP versus PP course (p=0.58) significantly affected the ARMSS score.

9.4.4 Definition of categories of alcohol consumption

Exposure to alcohol consumption was analyzed separately in men and women, as suggested (Baraona et al, 2001). Following the latest recommendation from the Italian Ministry for Health (<https://www.epicentro.iss.it/passi/rapporto2010/>), supported also by the European Centre for Disease Control and Prevention, we grouped patients into three categories (no consumption, moderate consumption, and high consumption), based on the reported number of weekly alcoholic units (AU) consumed (Table 9.4.1). An alcoholic unit corresponds to 12 g of alcohol, and it is approximately defined as 125 ml of wine, 333 ml of beer, 40 ml of spirits.

	<i>Male</i>	<i>Female</i>
<i>No consumption</i>	0 AU/week	0 AU/week
<i>Moderate consumption</i>	1-14 AU/week	1-7 AU/week
<i>High consumption</i>	>14 AU/week	>7 AU/week

Table 9.4.1 Classification of patients into categories for alcohol consumptions. AU = alcoholic unit.

9.5 Study 5: Genetic factors driving silent inflammation in progressive MS

9.5.1 Study cohorts for the genetic analysis

We compared patients with relapsing-remitting MS (RR-MS) versus secondary progressive MS (SP-MS). Patients in the RR-MS group had a confirmed relapsing-remitting course after at least 20 years from disease onset and a maximum Expanded Disability Status Scale (EDSS) score of 3.5 at last follow-up. Conversely, in patients of the SP-MS group, disease progression had been confirmed within 20 years from disease onset and EDSS was at least 4.0 at last follow-up (Table 7.1).

	<i>RR-MS</i>	<i>SP-MS</i>
Confirmed clinical course	=> 20 years from onset	< 20 years from onset
EDSS at latest follow-up	<= 3.5	=> 4.0

Table 9.5.1. Inclusion criteria for the patients involved in the genetic association study.

Starting from a large cohort of MS patients with available whole-genome genotype data at the Laboratory of Human Genetics of Neurological Disorders at OSR, we identified 772 patients with definite diagnosis of MS according to 2017 revised criteria (Thompson AJ et al, 2018), who fulfilled the inclusion criteria for RR-MS (n=406) or SP-MS (n=366) reported in Table 7.1.

For the SWE cohort, the same criteria were applied, leading to inclusion of a total of 2,062 patients (RR-MS = 863, SP-MS = 1199).

9.5.2 Genome-wide quality controls in the OSR cohort

Imputation to Haplotype Reference Consortium (HRC) reference genome and quality controls during the imputation process were performed as already described in Chapter 9.1. In total, starting of the 772 subjects, 13 did not pass the imputation QC.

On the remaining 759 subjects and 9,388,047 imputed genetic variants (reference genome: GRCh19/hg37), we applied a filter for missingness ≥ 0.02 (no subject was excluded). To discovery individuals with potential biological relationship, we performed a pairwise identity by descent (IBD) estimate, that identified 6 subjects potentially related ($PI_HAT \geq 0.250$). For each pair of subjects, the individuals who were the least characterized were removed (n=3). To identify population outliers (more than ± 6 standard deviations from population mean), we run a Principal Component Analysis (PCA) and excluded one subject. The final number of subjects included in the study was 755. For both IBD estimation and PCA, the regions with extended linkage disequilibrium (LD) were removed and the dataset was pruned (no pair of SNPs with $R^2 > 0.1$ in a window of 1000 kb).

Subsequently, we filtered out all the SNPs with $MAF \leq 0.05$, genotyping call rate $< 99\%$ and Hardy-Weinberg Equilibrium ($p < 1e-06$), for a total of 4,272,520 genetic variants in 755 subjects.

9.5.3 Selection of SNPs in iron metabolism genes

To determine a comprehensive list of genes involved in iron metabolism, we manually screened Gene Ontology (GO) and Human Phenotype (HP) terms related to iron homeostasis (Ashburner et al, 2000). This list of genes was merged with the Kyoto Encyclopedia of Genes and Genome (KEGG) ‘*Ferroptosis*’ pathway (Kanehisa et al,

2000). In total, we identified 336 genes relevant to iron metabolism. GRCh37/hg19 genes coordinates were extracted using UCSC Table Browser (Karolchik et al, 2004). 334 out of 336 genes were mapped, and the longest isoform was selected, adding a 2 kb flanking region to enrich for promoter and enhancer regions. An extensive list of mapped genes and pathways and gene mapping is reported in Appendix - Supplementary Tables S7.1 and S7.2.

Starting from whole genome data, we then extracted the mapped regions, resulting in 23,019 imputed variants and 755 subjects undergoing downstream association analysis. At this step, no filter for LD was applied to the variants.

9.5.4 Genetic association analysis in the discovery cohort

To calculate a threshold for statistical significance that was able to provide robust results despite multiple testing, we pruned the variants accounting for linkage disequilibrium ($R^2 \geq 0.6$). The number of genetically independent variants left was considered the total number of independent tests and a Bonferroni correction was applied to account for type I error (Benke et al, 2013). Following this approach, the threshold for significance was established at $p < 3.08e-05$ ($0.05/1,625$).

We then tested the association between genotypes for the selected variants and MS course, comparing RR-MS versus SP-MS using logistic regression and adjusting for sex and the first eight principal components to account for population stratification. PLINK v1.9 software was used for all the genetic QC and the association analysis (Chang et al, 2015).

9.5.5 Replication of the genetic association

The subjects of the replication cohort were part of the National Swedish MS Registry. Clinical information and blood samples for genotyping were gathered and managed at the Karolinska Institutet (Stockholm, Sweden). Reference genome for the imputation was GRCh37/hg19. Imputation and QC were run as described for the OSR cohort. A total of 2,062 patients with benign RRMS (n=863) and SPMS (n=1199) were included, who fulfilled the same criteria used for the discovery cohort and had available imputed genetic data. Sex and the first eight principal components were used as covariates.

9.5.6 RNA-seq and eQTL effect in PBMC

The subjects for this experiment were recruited at OSR. All patients were DMT-naive at the time of sampling, were not treated with corticosteroid drugs in the 30 days prior to sampling and had a diagnosis of RR-MS.

After blood sampling, Peripheral Blood Mononuclear Cells (PBMC) were isolated through density gradient using Lymphoprep (StemCell Technologies). RNA was extracted using Quick-DNA/RNA Miniprep Plus (Zymo), according to the manufacturing protocol. Quantity and quality of RNA was assessed by Qubit (Thermo Fisher Scientific) and TapeStation (Agilent). The transcriptomic profile of patients was obtained via next-generation sequencing technology. RNA libraries were generated using the TruSeq Stranded mRNA Library Prep Kit (Illumina) and sequenced on a HiSeq3000 sequencer (Illumina), reaching >25 million reads/samples. RNA-seq reads were aligned to hg19 reference genome by means of STAR tool (Dobin et al, 2013), using Trimmomatic (Bolger et al, 2014) to remove poor-quality bases and adapter sequences. Quantification of gene expression levels and gene-level summarization was performed using featureCounts (Liao et al, 2014), according to GENCODE v19 annotation gene model. Quality control of raw and aligned reads was performed by means of MultiQC tool (Ewels et al, 2016). To discard features deemed as not expressed, we only retained those with > 5 counts in at least 25% of the whole cohort.

The expression-Quantitative Trait Loci (eQTL) effect was computed via linear modelling, implying HIF1A normalized counts as outcome and the genotype (TT versus AT/AA) as predictor, adjusting for sex and age at sampling.

9.5.7 Plasma NFL

Plasma neurofilament light chain (pNFL) samples belonged to a previously published cohort of MS patients coming from a Swedish nationwide surveillance phase 4 study on DMT monitoring (Hillert et al, 2015). All patients had a diagnosis of MS and were naive or had been treated with glatiramer acetate or interferon and were sampled before the beginning of the first or new DMT and at regular intervals after treatment start. NFL levels were determined from frozen EDTA plasma samples using a sensitive immunoassay on the Simoa platform through a commercially available NF-Light kit and

antibodies from UmanDiagnostics according to the manufacturer's instructions (Quanterix, Lexington, MA).

All the pNFL levels have been adjusted for BMI and age at sampling, as suggested by recent literature (Benkert et al, 2022) and implied in linear modelling in form of a z-score, after rank normalization using the *RNOmni* R package to fit a normal distribution (McCaw et al, 2020). Normality of the data was then confirmed with the Shapiro-Wilk test. In addition, adjustment for disease duration at sampling and line of treatment about to start was also performed for pNFL in the experiment described in Chapter 7.3.3.

When evaluating the response to a new DMT in terms of change of pNFL, we built mixed linear models, adopting a paired-samples design.

9.5.8 Cerebrospinal fluid NFL

Cerebrospinal fluid (CSF) NFL samples were collected at the Karolinska Institutet (Stockholm, Sweden). After sampling, the CSF was centrifuged immediately and stored at -80°C. NFL levels were measured using commercially available ELISA kits (Uman diagnostics, Umeå, Sweden) according to the manufacturers' instructions. CSF NFL levels were normalized as described for pNFL. The analysis on CSF NFL in RR patients was adjusted for disease duration and age at sampling.

9.5.9 Pharmacogenomic study of DMF

To assess the contribution of the rs11621525 variant on the clinical response to treatment with Dimethyl Fumarate (DMF), we studied a cohort of 138 patients from OSR. Inclusion criteria were: 1) diagnosis of RR-MS; 2) not previously treated with second line DMT or other immunosuppressive drugs; 3) continuous DMF treatment for 2-years or shift to other DMT for lack of efficacy of DMF; 4) availability of clinical information to define the NEDA-3 status at two-years from DMF start; 5) availability of genotyping data.

9.5.10 Paramagnetic Rim Lesions (PRL) analysis

9.5.10.1 MRI acquisition

MRI acquisition was performed at the Neuroimaging of CNS White Matter Unit at OSR using a 3.0 Tesla Philips Ingenia CX scanner (Philips Medical Systems). The

following brain MRI sequences were acquired (receiving Coil=dS-Head-32): (a) sagittal three dimensional (3D) fluid attenuation inversion recovery (FLAIR), field of view (FOV)=256×256 mm, voxel size=1×1×1 mm, 192 slices, matrix=256×256, repetition time (TR)=4800 ms, echo time (TE)=270 ms, inversion time (TI)=1650 ms, echo train length (ETL)=167, acquisition time (TA)=6.15 min; (b) sagittal 3D T1-weighted turbo field echo, FOV=256×256, voxel size=1×1×1 mm, 204 slices, matrix=256×256, TR=7 ms, TE=3.2 ms, TI=1000 ms, flip angle=8°, TA=8.53 min; (c) 3D T2-weighted scan, FOV=256×256 mm, pixel size=1×1 mm, 192 axial slices with 1 mm slice thickness, TR 2500 ms, TE 330 ms, ETL=117, TA= 3 min; (c) 3D susceptibility-weighted image (SWI), FOV=230x230, pixel size=0.60×0.60 mm, 135 slices, 2 mm-thick, matrix=384x382, TR=39 ms, TEs=5.5:6:35.5 ms, flip angle=17°, TA=6 min; both magnitude and phase images for each echo were saved. For all scans, the slices were positioned parallel to a line joining the most infero-anterior and infero-posterior margins of the corpus callosum.

9.5.10.2 Conventional MRI analysis

Brain T2-hyperintense WM lesions were identified by a fully automated approach using the 3D FLAIR and 3D T1weighted as input images (Valverde et al, 2017). T2-hyperintense WM lesion volume was obtained for each patient from their lesion masks, after a careful visual check of the results provided by the automatic segmentation.

9.5.10.3 SWI processing

Maps of local B0 field changes were derived from the multi-echo SWIs using the package QSM (Sun and Wilman, 2015) for MatLab (The MathWorks Inc., Naticks, USA). Briefly, we first unwrapped phase images to eliminate discontinuities due to the limited range of phase values, using the best path method (Abdul-Rahman et al, 2007).

Subsequently, using a magnitude-weighted least square regression, we fit unwrapped phase images to the echo time. Finally, to make paramagnetic rim visible, we removed global spatial changes of the main magnetic field using regularization enabled sophisticated harmonic artefact reduction for phase data (Sun et al, 2014).

The 3D FLAIR image and the T2-hyperintense white matter lesion mask were then registered onto the SWI space using the magnitude of the first echo of the SWI sequence as reference image, which contains anatomical information, through rigid transformations

minimizing the normalized mutual information as cost function using FLIRT (FMRIB's Linear Image Registration Tool) embedded in FSL.

9.5.10.4 Quantification of paramagnetic rim lesion number and volume

Paramagnetic rim lesions (PRL) were defined as discrete FLAIR-hyperintense lesions either completely or partially surrounded by a paramagnetic rim of hypointense signal in unwrapped phase images (Figure 7.7).

For each patient, the number and volume of total T2-hyperintense WM lesions as well as of T2-hyperintense WM lesions with or without the hypointense paramagnetic rim were automatically estimated using MatLab.

Firstly, from the global lesion mask, different intensity values were manually given according to the different type of lesions (1=non-paramagnetic rim lesion, 2=paramagnetic rim lesion) creating a new label mask. Then, an automatic pipeline estimated the number and the dimension of the 3D connected objects (lesions) found within label masks, separately for each type of lesion (intensity value). Total number of T2-hyperintense WM lesions was obtained from the sum of T2-hyperintense WM lesions with or without the hypointense paramagnetic rim. Volumes were finally obtained by correcting the dimension for the voxel size.

9.5.10.5 PRL volume analysis

Starting from 103 patients with diagnosis of RR-MS or SP-MS, available MRI data on the PRL and available genetic data at OSR, a total of 47 patients (46%) had at least one PRL, while the remaining 56 had no PRL. We selected patients with maximum disease duration from onset of 20 years (n=35). The mean volumes were rank transformed to fit a normal distribution and the impact of the rs11621525 A allele on the lesion mean volume was assessed through linear modelling, adjusting for sex and total number of PRL.

References

- Abdul-Rahman HS, Gdeisat MA, Burton DR, Lalor MJ, Lilley F, Moore CJ. Fast and robust three-dimensional best path phase unwrapping algorithm. *Appl Opt*. 2007;46(26):6623-35.
- Abreu AP, Toro CA, Song YB, Navarro VM, Bosch MA, Eren A, Liang JN, Carroll RS, Latronico AC, Rønnekleiv OK, Aylwin CF, Lomniczi A, Ojeda S, Kaiser UB. MKRN3 inhibits the reproductive axis through actions in kisspeptin-expressing neurons. *J Clin Invest*. 2020 Aug 3;130(8):4486-4500.
- Absinta M, Sati P, Schindler M, Leibovitch EC, Ohayon J, Wu T, Meani A, Filippi M, Jacobson S, Cortese IC, Reich DS. Persistent 7-tesla phase rim predicts poor outcome in new multiple sclerosis patient lesions. *J Clin Invest*. 2016 Jul 1;126(7):2597-609.
- Absinta M, Sati P, Fechner A, Schindler MK, Nair G, Reich DS. Identification of chronic Active Multiple Sclerosis Lesions on 3T MRI. *AJNR Am J Neuroradiol*. 2018 Jul;39(7):1233-1238.
- Absinta M, Sati P, Masuzzo F, Nair G, Sethi V, Kolb H, Ohayon J, Wu T, Cortese ICM, Reich DS. Association of chronic Active Multiple Sclerosis Lesions With Disability In Vivo. *JAMA Neurol*. 2019 Dec 1;76(12):1474-1483.
- Absinta M, Maric D, Gharagozloo M, Garton T, Smith MD, Jin J, Fitzgerald KC, Song A, Liu P, Lin JP, Wu T, Johnson KR, McGavern DB, Schafer DP, Calabresi PA, Reich DS. A lymphocyte-microglia-astrocyte axis in chronic active multiple sclerosis. *Nature*. 2021 Sep;597(7878):709-714.
- Alamri A, Rahman R, Zhang M, Alamri A, Gounni AS, Kung SKP. Semaphorin-3E Produced by Immature Dendritic Cells Regulates Activated Natural Killer Cells Migration. *Front Immunol*. 2018 May 9;9:1005.
- Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA (2014). Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA Methylation microarrays. *Bioinformatics*, 30(10), 1363–1369
- Ascherio A, Munger KL, White R, Köchert K, Simon KC, Polman CH, Freedman MS, Hartung HP, Miller DH, Montalbán X, Edan G, Barkhof F, Pleimes D, Radü EW, Sandbrink R, Kappos L, Pohl C. Vitamin D as an early predictor of multiple sclerosis activity and progression. *JAMA Neurol*. 2014 Mar;71(3):306-14.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000 May;25(1):25-9.

Banwell B, Giovannoni G, Hawkes C, Lublin F. Editors' welcome and a working definition for a multiple sclerosis cure. *Mult Scler Relat Disord.* 2013 Apr;2(2):65-7.

Bar-Or A, Li R. Cellular immunology of relapsing multiple sclerosis: interactions, checks, and balances. *Lancet Neurol.* 2021 Jun;20(6):470-483.

Baranzini SE, Mudge J, Van Velkinburgh JC, Khankhanian P, Khrebtukova I, Miller NA, et al. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature.* 2010;464:1351–6

Baraona E, Abittan CS, Dohmen K, Moretti M, Pozzato G, Chayes ZW, Schaefer C, Lieber CS. Gender differences in pharmacokinetics of alcohol. *Alcohol Clin Exp Res.* 2001 Apr;25(4):502-7.

Benke KS, Wu Y, Fallin DM, Maher B, Palmer LJ. Strategy to control type I error increases power to identify genetic variation using the full biological trajectory. *Genet Epidemiol.* 2013 Jul;37(5):419-30.

Benkert P, Meier S, Schaedelin S, Manouchehrinia A, Yaldizli Ö, Maceski A, Oechtering J, Achtnichts L, Conen D, Derfuss T, Lalive PH, Mueller C, Müller S, Naegelin Y, Oksenberg JR, Pot C, Salmen A, Willemsse E, Kockum I, Blennow K, Zetterberg H, Gobbi C, Kappos L, Wiendl H, Berger K, Sormani MP, Granziera C, Piehl F, Leppert D, Kuhle J; NfL Reference Database in the Swiss Multiple Sclerosis Cohort Study Group. Serum neurofilament light chain for individual prognostication of disease activity in people with multiple sclerosis: a retrospective modelling and validation study. *Lancet Neurol.* 2022 Mar;21(3):246-257.

Berer K, Mues M, Koutrolos M, AlRasbi Z, Boziki M, Johner C, Wekerle H & Krishnamoorthy G (2011) Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination. *Nature* 479: 538–541

Berer K, Gerdes LA, Cekanaviciute E, Jia X, Xiao L, Xia Z, Liu C, Klotz L, Stauffer U, Baranzini SE, Kämpfel T, Hohlfeld R, Krishnamoorthy G, Wekerle H. (2017). Gut microbiota from multiple sclerosis patients enables spontaneous autoimmune

encephalomyelitis in mice. *Proceedings of the National Academy of Sciences of the United States of America*, 114(40), 10719–10724.

Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, Elledge SJ, Niebuhr DW, Scher AI, Munger KL, Ascherio A. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science*. 2022 Jan 21;375(6578):296-301.

Blair HA. Dimethyl Fumarate: A Review in Relapsing-Remitting MS. *Drugs*. 2019 Dec;79(18):1965-1976.

Blanco E, González-Ramírez M, Alcaine-Colet A, Aranda S, Di Croce L. The Bivalent Genome: Characterization, Structure, and Regulation. *Trends Genet*. 2020 Feb;36(2):118-131.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014 Aug 1;30(15):2114-20.

Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, van Iterson M, van Dijk F, van Galen M, Bot J, Sliker RC, Jhamai PM, Verbiest M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet*. 2017 Jan;49(1):131-138.

Bos SD, Page CM, Andreassen BK, Elboudwarej E, Gustavsen MW, Briggs F, et al. Genome-wide DNA methylation profiles indicate CD8⁺ T cell hypermethylation in multiple sclerosis. *PLoS One*. 2015;10:1–16.

Bove R, Chitnis T. The role of gender and sex hormones in determining the onset and outcome of multiple sclerosis. *Mult Scler*. 2014 Apr;20(5):520-6.

Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;44:512-25.

Bridel C, Leurs CE, van Lierop ZYGJ, van Kempen ZLE, Dekker I, Twaalfhoven HAM, Moraal B, Barkhof F, Uitdehaag BMJ, Killestein J, Teunissen CE. Serum Neurofilament Light Association With Progression in Natalizumab-Treated Patients With Relapsing-Remitting Multiple Sclerosis. *Neurology*. 2021 Nov 9;97(19):e1898-e1905.

Briggs KJ, Koivunen P, Cao S, Backus KM, Olenchock BA, Patel H, Zhang Q, Signoretti S, Gerfen GJ, Richardson AL, Witkiewicz AK, Cravatt BF, Clardy J, Kaelin

WG Jr. Paracrine Induction of HIF by Glutamate in Breast Cancer: EglN1 Senses Cysteine. *Cell*. 2016 Jun 30;166(1):126-39.

Brownlee WJ, Miszkiel KA, Tur C, Barkhof F, Miller DH, Ciccarelli O. Inclusion of optic nerve involvement in dissemination in space criteria for multiple sclerosis. *Neurology*. 2018 Sep 18;91(12):e1130-e1134.

Brynedal B, Duvefelt K, Jonasdottir G, Roos IM, Akesson E, Palmgren J, Hillert J. HLA-A confers an HLA-DRB1 independent influence on the risk of multiple sclerosis. *PLoS One*. 2007 Jul 25;2(7):e664.

Cagol A, Schaedelin S, Barakovic M, Benkert P, Todea RA, Rahmanzadeh R, Galbusera R, Lu PJ, Weigel M, Melie-Garcia L, Ruberte E, Siebenborn N, Battaglini M, Radue EW, Yaldizli Ö, Oechtering J, Sinnecker T, Lorscheider J, Fischer-Barnicol B, Müller S, Achtnichts L, Vehoff J, Disanto G, Findling O, Chan A, Salmen A, Pot C, Bridel C, Zecca C, Derfuss T, Lieb JM, Remonda L, Wagner F, Vargas MI, Du Pasquier R, Lalive PH, Pravata E, Weber J, Cattin PC, Gobbi C, Leppert D, Kappos L, Kuhle J, Granziera C. Association of Brain Atrophy With Disease Progression Independent of Relapse Activity in Patients With Relapsing Multiple Sclerosis. *JAMA Neurol*. 2022 Jul 1;79(7):682-692.

Campagna MP, Xavier A, Lea RA, Stankovich J, Maltby VE, Butzkueven H, Lechner-Scott J, Scott RJ, Jokubaitis VG. Whole-blood methylation signatures are associated with and accurately classify multiple sclerosis disease severity. *Clin Epigenetics*. 2022 Dec 30;14(1):194.

Carlström KE, Ewing E, Granqvist M, Gyllenberg A, Aeinehband S, Enoksson SL, Checa A, Badam TV, Huang J, Gomez-Cabrero D, Gustafsson M, Al Nimer F, Wheelock CE, Kockum I, Olsson T, Jagodic M, Piehl F. Therapeutic efficacy of dimethyl fumarate in relapsing-remitting multiple sclerosis associates with ROS pathway in monocytes. *Nat Commun*. 2019 Jul 12;10(1):3081.

Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4.

Chitnis T. Role of puberty in multiple sclerosis risk and course. *Clin Immunol*. 2013 Nov;149(2):192-200.

Clark IC, Gutiérrez-Vázquez C, Wheeler MA, Li Z, Rothhammer V, Linnerbauer M, Sanmarco LM, Guo L, Blain M, Zandee SEJ, Chao CC, Batterman KV, Schwabenland

M, Lotfy P, Tejeda-Velarde A, Hewson P, Manganeli Polonio C, Shultis MW, Salem Y, Tjon EC, Fonseca-Castro PH, Borucki DM, Alves de Lima K, Plasencia A, Abate AR, Rosene DL, Hodgetts KJ, Prinz M, Antel JP, Prat A, Quintana FJ. Barcoded viral tracing of single-cell interactions in central nervous system inflammation. *Science*. 2021 Apr 23;372(6540):eabf1230

Confavreux C, Vukusic S. Age at disability milestones in multiple sclerosis. *Brain*. 2006 Mar;129(Pt 3):595-605.

Conti L, De Palma R, Rolla S, Boselli D, Rodolico G, Kaur S, Silvennoinen O, Niccolai E, Amedei A, Ivaldi F, Clerico M, Contessa G, Uccelli A, Durelli L, Novelli F. Th17 cells in multiple sclerosis express higher levels of JAK2, which increases their surface expression of IFN- γ R2. *J Immunol*. 2012 Feb 1;188(3):1011-8.

Correale J, Hohlfeld, R, Baranzini SE. (2022). The role of the gut microbiota in multiple sclerosis. *Nature reviews. Neurology*, 18(9), 544–558.

Cosorich I, Dalla Costa G, Sorini C, Ferrarese R, Messina MJ, Dolpady J, Radice E, Mariani A, Testoni PA., Canducci F, Comi G, Martinelli V, Falcone M. (2017). High frequency of intestinal TH17 cells correlates with microbiota alterations and disease activity in multiple sclerosis. *Science advances*, 3(7), e1700492.

Costa C, Martínez-Sáez E, Gutiérrez-Franco A, Eixarch H, Castro Z, Ortega-Aznar A, Ramón Y Cajal S, Montalban X, Espejo C. Expression of semaphorin 3A, semaphorin 7A and their receptors in multiple sclerosis lesions. *Mult Scler*. 2015 Nov;21(13):1632-43.

Cox LM, Maghzi AH, Liu S, Tankou SK, Dhang FH, Willocq V, Song A, Wasén C, Tauhid S, Chu R, Anderson MC, De Jager PL, Polgar-Turcsanyi M, Healy BC, Glanz BI, Bakshi R, Chitnis T, Weiner HL. Gut Microbiome in Progressive Multiple Sclerosis. *Ann Neurol*. 2021 Jun;89(6):1195-1211.

Cramer T, Yamanishi Y, Clausen BE, Förster I, Pawlinski R, Mackman N, Haase VH, Jaenisch R, Corr M, Nizet V, Firestein GS, Gerber HP, Ferrara N, Johnson RS. HIF-1 α is essential for myeloid cell-mediated inflammation. *Cell*. 2003 Mar 7;112(5):645-57.

Cyran AM, Zhitkovich A. HIF1, HSF1, and NRF2: Oxidant-Responsive Trio Raising Cellular Defenses and Engaging Immune System. *Chem Res Toxicol*. 2022 Oct 17;35(10):1690-1700.

D'hooghe MB, Haentjens P, Nagels G, D'Hooghe T, De Keyser J. Menarche, oral contraceptives, pregnancy and progression of disability in relapsing onset and progressive onset multiple sclerosis. *J Neurol*. 2012 May;259(5):855-61.

Dang EV, Barbi J, Yang HY, Jinasena D, Yu H, Zheng Y, Bordman Z, Fu J, Kim Y, Yen HR, Luo W, Zeller K, Shimoda L, Topalian SL, Semenza GL, Dang CV, Pardoll DM, Pan F. Control of T(H)17/T(reg) balance by hypoxia-inducible factor 1. *Cell*. 2011 Sep 2;146(5):772-84.

De Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, Ke X, Monsuur AJ, Whittaker P, Delgado M, Morrison J, Richardson A, Walsh EC, Gao X, Galver L, Hart J, Hafler DA, Pericak-Vance M, Todd JA, Daly MJ, Trowsdale J, Wijmenga C, Vyse TJ, Beck S, Murray SS, Carrington M, Gregory S, Deloukas P, Rioux JD. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet*. 2006 Oct;38(10):1166-72.

De Flon P, Gunnarsson M, Laurell K, Söderström L, Birgander R, Lindqvist T, Krauss W, Dring A, Bergman J, Sundström P, Svenningsson A. Reduced inflammation in relapsing-remitting multiple sclerosis after therapy switch to rituximab. *Neurology*. 2016 Jul 12;87(2):141-7.

Dewailly D, Andersen CY, Balen A, Broekmans F, Dilaver N, Fanchin R, Griesinger G, Kelsey TW, La Marca A, Lambalk C, Mason H, Nelson SM, Visser JA, Wallace WH, Anderson RA. The physiology and clinical utility of anti-Mullerian hormone in women. *Hum Reprod Update*. 2014 May-Jun;20(3):370-85.

Díaz C, Zarco LA, Rivera DM. Highly active multiple sclerosis: An update. *Mult Scler Relat Disord*. 2019 May;30:215-224.

Diaz-Cruz C, Chua AS, Malik MT, Kaplan T, Glanz BI, Egorova S, Guttmann CRG, Bakshi R, Weiner HL, Healy BC, Chitnis T. The effect of alcohol and red wine consumption on clinical and MRI outcomes in multiple sclerosis. *Mult Scler Relat Disord*. 2017 Oct;17:47-53.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013 Jan 1;29(1):15-21.

Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010 Nov 30;11:587.

Dugger SA, Platt A, Goldstein DB. Drug development in the era of precision medicine. *Nat Rev Drug Discov*. 2018 Mar;17(3):183-196.

Eiza N, Garty M, Staun-Ram E, Miller A, Vadasz Z. The possible involvement of sema3A and sema4A in the pathogenesis of multiple sclerosis. *Clin Immunol*. 2022 May;238:109017.

Elliott C, Wolinsky JS, Hauser SL, Kappos L, Barkhof F, Bernasconi C, Wei W, Belachew S, Arnold DL. Slowly expanding/evolving lesions as a magnetic resonance imaging marker of chronic active multiple sclerosis lesions. *Mult Scler*. 2019 Dec;25(14):1915-1925.

Eshaghi A, Marinescu RV, Young AL, Firth NC, Prados F, Jorge Cardoso M, Tur C, De Angelis F, Cawley N, Brownlee WJ, De Stefano N, Laura Stromillo M, Battaglini M, Ruggieri S, Gasperini C, Filippi M, Rocca MA, Rovira A, Sastre-Garriga J, Geurts JJG, Vrenken H, Wottschel V, Leurs CE, Uitdehaag B, Pirpamer L, Enzinger C, Ourselin S, Gandini Wheeler-Kingshott CA, Chard D, Thompson AJ, Barkhof F, Alexander DC, Ciccarelli O. Progression of regional grey matter atrophy in multiple sclerosis. *Brain*. 2018 Jun 1;141(6):1665-1677.

Ewels P, Magnusson M, Lundin S, Källér M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016 Oct 1;32(19):3047-8.

Fahim M, Rafiee Zadeh A, Shoureshi P, Ghadimi K, Cheshmavar M, Sheikhinia N, Afzali M. Alcohol and multiple sclerosis: an immune system-based review. *Int J Physiol Pathophysiol Pharmacol*. 2020 Apr 15;12(2):58-69.

Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, Jostins L, Plant K, Andrews R, McGee C, Knight JC. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*. 2014 Mar 7;343(6175):1246949.

Feinstein A, Freeman J, Lo AC. Treatment of progressive multiple sclerosis: what works, what does not, and what is needed. *Lancet Neurol*. 2015 Feb;14(2):194-207

Filippi M, Bar-Or A, Piehl F, Preziosa P, Solari A, Vukusic S, Rocca MA. Multiple sclerosis. *Nat Rev Dis Primers*. 2018 Nov 8;4(1):43

Filippi M, Preziosa P, Banwell BL, Barkhof F, Ciccarelli O, De Stefano N, Geurts JGG, Paul F, Reich DS, Toosy AT, Traboulsee A, Wattjes MP, Yousry TA, Gass A, Lubetzki C, Weinshenker BG, Rocca MA. Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines. *Brain*. 2019 Jul 1;142(7):1858-1875.

Fisniku LK, Brex PA, Altmann DR, et al. Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain* 2008; 131: 808– 17.

Fitzgerald KC, Munger KL, Köchert K, Arnason BG, Comi G, Cook S, Goodin DS, Filippi M, Hartung HP, Jeffery DR, O'Connor P, Suarez G, Sandbrink R, Kappos L, Pohl C, Ascherio A. Association of Vitamin D Levels With Multiple Sclerosis Activity and Progression in Patients Receiving Interferon Beta-1b. *JAMA Neurol*. 2015 Dec;72(12):1458-65.

Galli E, Hartmann FJ, Schreiner B, Ingelfinger F, Arvaniti E, Diebold M, Mrdjen D, van der Meer F, Krieg C, Nimer FA, Sanderson N, Stadelmann C, Khademi M, Piehl F, Claassen M, Derfuss T, Olsson T, Becher B. GM-CSF and CXCR4 define a T helper cell signature in multiple sclerosis. *Nat Med*. 2019 Aug;25(8):1290-1300.

Gava G, Bartolomei I, Costantino A, Berra M, Venturoli S, Salvi F, Meriggiola MC. Long-term influence of combined oral contraceptive use on the clinical course of relapsing-remitting multiple sclerosis. *Fertil Steril*. 2014 Jul;102(1):116-22.

Ghadirian, P., Dadgostar, B., Azani, R., Maisonneuve, P. A case-control study of the association between socio-demographic, lifestyle and medical history factors and multiple sclerosis. *Can. J. Public Health* 92, 281–285 (2001).

Ghousaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, Fumis L, Miranda A, Carvalho-Silva D, Buniello A, Burdett T, Hayhurst J, Baker J, Ferrer J, Gonzalez-Uriarte A, Jupp S, Karim MA, Koscielny G, Machlitt-Northen S, Malangone C, Pendlington ZM, Roncaglia P, Suveges D, Wright D, Vrousseau O, Papa E, Parkinson H, MacArthur JAL, Todd JA, Barrett JC, Schwartzenuber J, Hulcoop DG, Ochoa D, McDonagh EM, Dunham I. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res*. 2021 Jan 8;49(D1):D1311-D1320.

Gianfrancesco MA, Stridh P, Rhead B, Shao X, Xu E, Graves JS, Chitnis T, Waldman A, Lotze T, Seiner T, Belman A, Greenberg B, Weinstock-Guttman B, Aaen G, Tillema JM, Hart J, Caillier S, Ness J, Harris Y, Rubin J, Candee M, Krupp L, Gorman M, Benson

L, Rodriguez M, Mar S, Kahn I, Rose J, Roalstad S, Casper TC, Shen L, Quach H, Quach D, Hillert J, Bäärnhielm M, Hedstrom A, Olsson T, Kockum I, Alfredsson L, Metayer C, Schaefer C, Barcellos LF, Waubant E; Network of Pediatric Multiple Sclerosis Centers. Evidence for a causal relationship between low vitamin D, high BMI, and pediatric-onset MS. *Neurology*. 2017 Apr 25;88(17):1623-1629.

Gombash SE, Lee PW, Sawdai E, Lovett-Racke AE. Vitamin D as a Risk Factor for Multiple Sclerosis: Immunoregulatory or Neuroprotective? *Front Neurol*. 2022 May 16;13:796933.

Goris A, Vandebergh M, McCauley JL, Saarela J, Cotsapas C. Genetics of multiple sclerosis: lessons from polygenicity. *Lancet Neurol*. 2022 Sep;21(9):830-842.

Graves JS, Henry RG, Cree BAC, Lambert-Messerlian G, Greenblatt RM, Waubant E, Cedars MI, Zhu A; University of California, San Francisco MS-EPIC Team,; Bacchetti P, Hauser SL, Oksenberg JR. Ovarian aging is associated with gray matter volume and disability in women with MS. *Neurology*. 2018 Jan 16;90(3):e254-e260.

Hara T, Nakaoka HJ, Hayashi T, Mimura K, Hoshino D, Inoue M, Nagamura F, Murakami Y, Seiki M, Sakamoto T. Control of metastatic niche formation by targeting APBA3/Mint3 in inflammatory monocytes. *Proc Natl Acad Sci U S A*. 2017 May 30;114(22):E4416-E4424.

Harding K, Williams O, Willis M, Hrastelj J, Rimmer A, Joseph F, Tomassini V, Wardle M, Pickersgill T, Robertson N, Tallantyre E. Clinical Outcomes of Escalation vs Early Intensive Disease-Modifying Therapy in Patients With Multiple Sclerosis. *JAMA Neurol*. 2019 May 1;76(5):536-541.

Harding KE, Robertson NP. New rare genetic variants in multiple sclerosis. *J Neurol*. 2019 Jan;266(1):278-280.

Hassani A, Corboy JR, Al-Salam S, Khan G. Epstein-Barr virus is present in the brain of most cases of multiple sclerosis and may engage more than just B cells. *PloS One* (2018) 13:e0192109. doi: 10.1371/journal.pone.0192109

Hausler D, Torke S, Peelen E, Bertsch T, Djukic M, Nau R, et al. High dose vitamin D exacerbates central nervous system autoimmunity by raising T-cell excitatory calcium. *Brain*. (2019) 142:2737–55.

Healy BC, Ali EN, Guttmann CR, Chitnis T, Glanz BI, Buckle G, Houtchens M, Stazzone L, Moodie J, Berger AM, Duan Y, Bakshi R, Khoury S, Weiner H, Ascherio A.

Smoking and disease progression in multiple sclerosis. *Arch Neurol.* 2009 Jul;66(7):858-64.

Hedström AK, Bäärnhielm M, Olsson T, Alfredsson L. Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. *Neurology.* 2009 Sep 1;73(9):696-701.

Hedström AK, Sundqvist E, Bäärnhielm M, Nordin N, Hillert J, Kockum I, Olsson T, Alfredsson L. Smoking and two human leukocyte antigen genes interact to increase the risk for multiple sclerosis. *Brain.* 2011 Mar;134(Pt 3):653-64.

Hedström AK, Bäärnhielm M, Olsson T, Alfredsson L. Exposure to environmental tobacco smoke is associated with increased risk for multiple sclerosis. *Mult Scler.* 2011 Jul;17(7):788-93.

Hedström AK, Lima Bomfim I, Barcellos L, Gianfrancesco M, Schaefer C, Kockum I, Olsson T, Alfredsson L. Interaction between adolescent obesity and HLA risk genes in the etiology of multiple sclerosis. *Neurology.* 2014a Mar 11;82(10):865-72.

Hedström AK, Hillert J, Olsson T & Alfredsson L. Alcohol as a modifiable lifestyle factor affecting multiple sclerosis risk. *JAMA Neurol.* 71, 300–305 (2014b).

Hedström AK, Olsson T, Alfredsson L. Body mass index during adolescence, rather than childhood, is critical in determining MS risk. *Mult Scler.* 2016 Jun;22(7):878-83.

Heinz R and Waltenbaugh C. Ethanol consumption modifies dendritic cell antigen presentation in mice. *Alcohol Clin Exp Res* 2007; 31: 1759-1771.

Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC. The MR-Base platform supports systematic causal inference across the human phenome. *Elife.* 2018 May 30;7:e34408.

Hemmer B, Kerschensteiner M, Korn T. Role of the innate and adaptive immune responses in the course of multiple sclerosis. *Lancet Neurol.* 2015 Apr;14(4):406-19.

Hillert J, Stawiarz L. The Swedish MS registry – clinical support tool and scientific resource. *Acta Neurol Scand.* 2015;132(199):11-9.

Hilven K, Vandebergh M, Smets I, Mallants K, Goris A, Dubois B. Genetic basis for relapse rate in multiple sclerosis: Association with LRP2 genetic variation. *Mult Scler.* 2018 Nov;24(13):1773-1775. doi: 10.1177/1352458517749894.

Houen G, Trier NH, Frederiksen JL. Epstein-Barr Virus and Multiple Sclerosis. *Front Immunol*. 2020 Dec 17;11:587078.

Houseman EA, Kile ML, Istiani DC, Ince TA, Kelsey KT, Marsit CJ. Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. *BMC Bioinformatics*. 2016 Jun 29;17:259. doi: 10.1186/s12859-016-1140-4.

Hupperts R, Smolders J, Vieth R, Holmøy T, Marhardt K, Schluep M, Killestein J, Barkhof F, Beelke M, Grimaldi LME; SOLAR Study Group. Randomized trial of daily high-dose vitamin D3 in patients with RRMS receiving subcutaneous interferon β -1a. *Neurology*. 2019 Nov 12;93(20):e1906-e1916.

Huynh JL, Garg P, Thin TH, Yoo S, Dutta R, Trapp BD, Haroutunian V, Zhu J, Donovan MJ, Sharp AJ, Casaccia P. Epigenome-wide differences in pathology-free regions of multiple sclerosis-affected brains. *Nat Neurosci*. 2014 Jan;17(1):121-30.

International Multiple Sclerosis Genetics Consortium; Wellcome Trust Case Control Consortium; Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*. 2011 Aug 10;476(7359):214-9.

International Multiple Sclerosis Genetics Consortium. Low-frequency and rare-coding variation contributes to multiple sclerosis risk. *Cell* 2018; 175: 1679–87.e7.

International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science*. 2019 Sep 27;365(6460):eaav7188.

Isaacs JS, Jung YJ, Mole DR, Lee S, Torres-Cabala C, Chung YL, Merino M, Trepel J, Zbar B, Toro J, Ratcliffe PJ, Linehan WM, Neckers L. HIF overexpression correlates with biallelic loss of fumarate hydratase in renal cancer: novel role of fumarate in regulation of HIF stability. *Cancer Cell*. 2005 Aug;8(2):143-53.

Ivashynka A, Copetti M, Naldi P, D'Alfonso S, Leone MA. The Impact of Lifetime Alcohol and Cigarette Smoking Loads on Multiple Sclerosis Severity. *Front Neurol*. 2019 Aug 13;10:866.

Jäkel S, Agirre E, Mendanha Falcão A, van Bruggen D, Lee KW, Knuesel I, Malhotra D, French-Constant C, Williams A, Castelo-Branco G. Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature*. 2019 Feb;566(7745):543-547.

Johnston CJ, Smyth DJ, Dresser DW, Maizels RM. TGF- β in tolerance, development and regulation of immunity. *Cell Immunol.* 2016 Jan;299:14-22.

Jongen PJ. Health-Related Quality of Life in Patients with Multiple Sclerosis: Impact of Disease-Modifying Drugs. *CNS Drugs.* 2017 Jul;31(7):585-602.

Källberg H, Jacobsen S, Bengtsson C, Pedersen M, Padyukov L, Garred P, Frisch M, Karlson EW, Klareskog L, Alfredsson L. Alcohol consumption is associated with decreased risk of rheumatoid arthritis: results from two Scandinavian case-control studies. *Ann Rheum Dis.* 2009 Feb;68(2):222-7.

Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000 Jan 1;28(1):27-30.

Kappos L, Radue E-W, O'Connor P, Polman C, Hohlfeld R, Calabresi P, Selmaj K, Agoropoulou C, Leyk M, Zhang-Auberson L, et al (2010) A Placebo-Controlled Trial of Oral Fingolimod in Relapsing Multiple Sclerosis. *N Engl J Med* 362: 387– 401

Kappos L, De Stefano N, Freedman MS, Cree BA, Radue EW, Sprenger T, Sormani MP, Smith T, Häring DA, Piani Meier D, Tomic D. Inclusion of brain volume loss in a revised measure of 'no evidence of disease activity' (NEDA-4) in relapsing-remitting multiple sclerosis. *Mult Scler.* 2016 Sep;22(10):1297-305.

Kappos L, Fox RJ, Burcklen M, Freedman MS, Havrdová EK, Hennessy B, Hohlfeld R, Lublin F, Montalban X, Pozzilli C, Scherz T, D'Ambrosio D, Linscheid P, Vaclavkova A, Pirozek-Lawniczek M, Kracker H, Sprenger T. Ponesimod Compared With Teriflunomide in Patients With Relapsing Multiple Sclerosis in the Active-Comparator Phase 3 OPTIMUM Study: A Randomized Clinical Trial. *JAMA Neurol.* 2021 May 1;78(5):558-567.

Karczewski, K.J., Francioli, L.C., Tiao, G. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443 (2020).

Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 2004 Jan 1;32(Database issue):D493-6

Kaufmann M, Schaupp AL, Sun R, Coscia F, Dendrou CA, Cortes A, Kaur G, Evans HG, Mollbrink A, Navarro JF, Sonner JK, Mayer C, DeLuca GC, Lundeberg J, Matthews PM, Attfield KE, Friese MA, Mann M, Fugger L. Identification of early

neurodegenerative pathways in progressive multiple sclerosis. *Nat Neurosci.* 2022 Jul;25(7):944-955.

Kister I, Kantarci OH. Multiple Sclerosis Severity Score: Concept and applications. *Mult Scler.* 2020 Apr;26(5):548-553.

Klutstein M, Nejman D, Greenfield R, Cedar H. DNA Methylation in Cancer and Aging. *Cancer Res.* 2016 Jun 15;76(12):3446-50.

Koch M, Kingwell E, Rieckmann P, Tremlett H. The natural history of primary progressive multiple sclerosis. *Neurology.* 2009 Dec 8;73(23):1996-2002.

Köhler T, Reizis B, Johnson RS, Weighardt H, Förster I. Influence of hypoxia-inducible factor 1 α on dendritic cell differentiation and migration. *Eur J Immunol.* 2012 May;42(5):1226-36.

Kuhle J, Barro C, Andreasson U, Derfuss T, Lindberg R, Sandelius Å, Liman V, Norgren N, Blennow K, Zetterberg H. Comparison of three analytical platforms for quantification of the neurofilament light chain in blood samples: ELISA, electrochemiluminescence immunoassay and Simoa. *Clin Chem Lab Med.* 2016 Oct 1;54(10):1655-61

Kuhle J, Plavina T, Barro C, Disanto G, Sangurdekar D, Singh CM, de Moor C, Engle B, Kieseier BC, Fisher E, Kappos L, Rudick RA, Goyal J. Neurofilament light levels are associated with long-term outcomes in multiple sclerosis. *Mult Scler.* 2020 Nov;26(13):1691-1699.

Kuhlmann T, Ludwin S, Prat A, Antel J, Brück W, Lassmann H. An updated histological classification system for multiple sclerosis lesions. *Acta Neuropathol.* 2017 Jan;133(1):13-24.

Kulakova OG, Kabilov MR, Danilova LV, Popova EV, Baturina OA, Tsareva EY, et al. Whole-genome DNA methylation analysis of peripheral blood mononuclear cells in multiple sclerosis patients with different disease courses. *Acta Naturae.* 2016;8:103–10.

Kular L, Liu Y, Ruhrmann S, Zheleznyakova G, Marabita F, Gomez-Cabrero D, James T, Ewing E, Lindén M, Górnikiewicz B, Aeinehband S, Stridh P, Link J, Andlauer TFM, Gasperi C, Wiendl H, Zipp F, Gold R, Tackenberg B, Weber F, Hemmer B, Strauch K, Heilmann-Heimbach S, Rawal R, Schminke U, Schmidt CO, Kacprowski T, Franke A, Laudes M, Dilthey AT, Celius EG, Søndergaard HB, Tegnér J, Harbo HF, Oturai AB, Olafsson S, Eggertsson HP, Halldorsson BV, Hjaltason H, Olafsson E, Jonsdottir I,

Stefansson K, Olsson T, Piehl F, Ekström TJ, Kockum I, Feinberg AP, Jagodic M. DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis. *Nat Commun.* 2018 Jun 19;9(1):2397.

Kular L, Jagodic M. Epigenetic insights into multiple sclerosis disease progression. *J Intern Med.* 2020 Jul;288(1):82-102.

Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology.* 1983 Nov;33(11):1444-52.

Langer-Gould A, Brara SM, Beaber BE, Koebnick C. Childhood obesity and risk of pediatric multiple sclerosis and clinically isolated syndrome. *Neurology.* 2013 Feb 5;80(6):548-52.

Lassmann H, van Horssen J, Mahad D. Progressive multiple sclerosis: pathology and pathogenesis. *Nat Rev Neurol.* 2012 Nov 5;8(11):647-56.

Lee YK, Mazmanian SK. Has the microbiota played a critical role in the evolution of the adaptive immune system?. *Science* 2010, 330(6012), 1768–1773

Leek JT, Johnson WE, Parker HS, Fertig EJ, Jaffe AE, Zhang Y, Storey JD, Torres LC (2022). *sva: Surrogate Variable Analysis*. R package version 3.46.0.

Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet.* 2010 Oct;11(10):733-9.

Li C, Lu W, Yang L, Li Z, Zhou X, Guo R, Wang J, Wu Z, Dong Z, Ning G, Shi Y, Gu Y, Chen P, Hao Z, Han T, Yang M, Wang W, Huang X, Li Y, Gao S, Hu R. MKRN3 regulates the epigenetic switch of mammalian puberty via ubiquitination of MBD3. *Natl Sci Rev.* 2020 Mar;7(3):671-685.

Li X, Zhang Y, Yan Y, Ciric B, Ma CG, Chin J, Curtis M, Rostami A, Zhang GX. LINGO-1-Fc-Transduced Neural Stem Cells Are Effective Therapy for chronic Stage Experimental Autoimmune Encephalomyelitis. *Mol Neurobiol.* 2017 Aug;54(6):4365-4378. doi: 10.1007/s12035-016-9994-z. Epub 2016 Jun 25.

Liao Y, Smyth GK, Shi W. *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* *Bioinformatics.* 2014 Apr 1;30(7):923-30.

Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. *WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs.* *Nucleic Acids Res.* 2019 Jul 2;47(W1):W199-W205.

Limesurvey GmbH. LimeSurvey: An Open Source survey tool. LimeSurvey GmbH, Hamburg, Germany. URL: <http://www.limesurvey.org>

Liu R, Du S, Zhao L, Jain S, Sahay K, Rizvanov A, Lezhnyova V, Khaibullin T, Martynova E, Khaiboullina S, Baranwal M. Autoreactive lymphocytes in multiple sclerosis: Pathogenesis and treatment target. *Front Immunol.* 2022 Sep 23;13:996469.

Løken-Amsrud KI, Holmøy T, Bakke SJ, Beiske AG, Bjerve KS, Bjørnarå BT, Hovdal H, Lilleås F, Midgard R, Pedersen T, Benth JS, Sandvik L, Torkildsen O, Wergeland S, Myhr KM. Vitamin D and disease activity in multiple sclerosis before and during interferon- β treatment. *Neurology.* 2012 Jul 17;79(3):267-73.

Luchtman D, Gollan R, Ellwardt E, Birkenstock J, Robohm K, Siffrin V, Zipp F. In vivo and in vitro effects of multiple sclerosis immunomodulatory therapeutics on glutamatergic excitotoxicity. *J Neurochem.* 2016 Mar;136(5):971-80.

Maggi P, Kuhle J, Schädelin S, van der Meer F, Weigel M, Galbusera R, Mathias A, Lu PJ, Rahmzadeh R, Benkert P, La Rosa F, Bach Cuadra M, Sati P, Théaudin M, Pot C, van Pesch V, Leppert D, Stadelmann C, Kappos L, Du Pasquier R, Reich DS, Absinta M, Granziera C. chronic White Matter Inflammation and Serum Neurofilament Levels in Multiple Sclerosis. *Neurology.* 2021 Aug 10;97(6):e543-e553.

Maggi P, Vanden Bulcke C, Pedrini E, Bugli C, Sellimi A, Wynen A, Ben Ayad A, Mullins WA, Kalaitzidis G, Lolli V, Perrotta G, El Sankari S, Duprez T, Li X, Calabresi PA, van Pesch V, Reich DS, Absinta M. Chronic active multiple sclerosis lesions are poorly responsive to anti-CD20 antibody treatment. Oral communication (O118). ECTRIMS 2022. *Mult Scler.* 2022 Oct; 28 (issue 3, suppl).

Maltby VE, Graves MC, Lea RA, Benton MC, Sanders KA, Tajouri L, et al. Genome-wide DNA methylation profiling of CD8⁺ T cells shows a distinct epigenetic signature to CD4⁺ T cells in multiple sclerosis patients. *Clin Epigenetics.* 2015;7:1–6.

Maltby VE, Lea RA, Graves MC, Sanders KA, Benton MC, Tajouri L, Scott RJ, Lechner-Scott J. Genome-wide DNA methylation changes in CD19⁺ B cells from relapsing-remitting multiple sclerosis patients. *Sci Rep.* 2018 Nov 27;8(1):17418.

Manouchehrinia A, Tench CR, Maxted J, Bibani RH, Britton J, Constantinescu CS. Tobacco smoking and disability progression in multiple sclerosis: United Kingdom cohort study. *Brain.* 2013 Jul;136(Pt 7):2298-304.

Manouchehrinia A, Westerlind H, Kingwell E, Zhu F, Carruthers R, Ramanujam R, Ban M, Glaser A, Sawcer S, Tremlett H, Hillert J. Age Related Multiple Sclerosis Severity Score: Disability ranked by age. *Mult Scler*. 2017 Dec;23(14):1938-1946.

Marabita F, Almgren M, Sjöholm LK, Kular L, Liu Y, James T, et al. Smoking induces DNA methylation changes in multiple sclerosis patients with exposure-response relationship. *Sci Rep*. 2017;7:1–15.

Martinelli V, Dalla Costa G, Colombo B, Dalla Libera D, Rubinacci A, Filippi M, Furlan R, Comi G. Vitamin D levels and risk of multiple sclerosis in patients with clinically isolated syndromes. *Mult Scler*. 2014 Feb;20(2):147-55.

Martire MS, Moiola L, Rocca MA, Filippi M, Absinta M. What is the potential of paramagnetic rim lesions as diagnostic indicators in multiple sclerosis? *Expert Rev Neurother*. 2022 Oct;22(10):829-837.

Mattei AL, Bailly N, Meissner A. DNA methylation: a historical perspective. *Trends Genet*. 2022 Jul;38(7):676-707.

McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, et al; Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016 Oct;48(10):1279-83.

McCaw ZR, Lane JM, Saxena R, Redline S, Lin X. Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics*. 2020 Dec;76(4):1262-1272.

Merad M, Sathe P, Helft J, Miller J, Mortha A. The dendritic cell lineage: ontogeny and function of dendritic cells and their subsets in the steady state and the inflamed setting. *Annu Rev Immunol*. 2013;31:563–604.

Meyer T, Shimon D, Youssef S, Yankovitz G, Tessler A, Chernobylsky T, Gaoni-Yogev A, Perelroizen R, Budick-Harmelin N, Steinman L, Mayo L. NAD⁺ metabolism drives astrocyte proinflammatory reprogramming in central nervous system autoimmunity. *Proc Natl Acad Sci U S A*. 2022 Aug 30;119(35):e2211310119. doi: 10.1073/pnas.2211310119. Epub 2022 Aug 22.

Meyer-Moock S, Feng YS, Maeurer M, Dippel FW, Kohlmann T. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS)

and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. *BMC Neurol.* 2014 Mar 25;14:58.

Mills EA, Ogrodnik MA, Plave A, Mao-Draayer Y. Emerging Understanding of the Mechanism of Action for Dimethyl Fumarate in the Treatment of Multiple Sclerosis. *Front Neurol.* 2018 Jan 23;9:5.

Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, Carnero-Montoro E, Lawson DJ, Burrows K, Suderman M, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet.* 2021 Sep;53(9):1311-1321.

Mirza, A., & Mao-Draayer, Y. (2017). The gut microbiome and microbial translocation in multiple sclerosis. *Clinical immunology*, 183, 213–224.

Mokry LE, Ross S, Timpson NJ, Sawcer S, Davey Smith G, Richards JB. Obesity and Multiple Sclerosis: A Mendelian Randomization Study. *PLoS Med.* 2016;13(6):e1002053. Published 2016 Jun 28.

Montalban X, Hauser SL, Kappos L, Arnold DL, Bar-Or A, Comi G, de Seze J, Giovannoni G, Hartung H-P, Hemmer B, et al (2017) Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis. *N Engl J Med* 376: 209–220

Moolhuijsen LME, Visser JA. Anti-Müllerian Hormone and Ovarian Reserve: Update on Assessing Ovarian Function. *J Clin Endocrinol Metab.* 2020 Nov 1;105(11):3361–73.

Mountjoy E, Schmidt EM, Carmona M, Schwartzentruber J, Peat G, Miranda A, Fumis L, Hayhurst J, Buniello A, Karim MA, Wright D, Hercules A, Papa E, Fauman EB, Barrett JC, Todd JA, Ochoa D, Dunham I, Ghousaini M. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat Genet.* 2021 Nov;53(11):1527-1533

Movassagh H, Koussih L, Shan L, Gounni AS. The regulatory role of semaphorin 3E in allergic asthma. *Int J Biochem Cell Biol.* 2019 Jan;106:68-73.

Munger K, Äivo J, Hongell K, Soilu-Hänninen M, Surcel H, Ascherio A. Vitamin D status during pregnancy and risk of multiple sclerosis in offspring of women in the Finnish maternity cohort. *JAMA Neurol.* 2016;73(5):515.

Munger K, Levin L, Hollis B, Howard N, Ascherio A. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. *JAMA.* 2006;296(23):2832.

Munger KL, Bentzen J, Laursen B, Stenager E, Koch-Henriksen N, Sørensen TI, Baker JL. Childhood body mass index and multiple sclerosis risk: a long-term cohort study. *Mult Scler.* 2013 Sep;19(10):1323-9.

Nachliely M, Trachtenberg A, Khalfin B, Nalbandyan K, Cohen-Lahav M, Yasuda K, et al. Dimethyl fumarate and vitamin D derivatives cooperatively enhance VDR and Nrf2 signaling in differentiating AML cells in vitro and inhibit leukemia progression in a xenograft mouse model. *J Steroid Biochem Mol Biol.* (2019) 188:8–16.

Ness KJ, Fan J, Wilke WW, Coleman RA, Cook RT and Schlueter AJ. Chronic ethanol consumption decreases murine Langerhans cell numbers and delays migration of Langerhans cells as well as dermal dendritic cells. *Alcohol Clin Exp Res* 2008; 32: 657-668.

Nicholas R, Straube S, Schmidli H, Pfeiffer S, Friede T. Time-patterns of annualized relapse rates in randomized placebo-controlled clinical trials in relapsing multiple sclerosis: a systematic review and meta-analysis. *Mult Scler.* 2012 Sep;18(9):1290-6.

Nizri E, Irony-Tur-Sinai M, Lory O, Orr-Urtreger A, Lavi E, Brenner T. Activation of the cholinergic anti-inflammatory system by nicotine attenuates neuroinflammation via suppression of Th1 and Th17 responses. *J Immunol.* 2009 Nov 15;183(10):6681-8.

Olsson T, Barcellos LF, Alfredsson L. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. *Nat Rev Neurol.* 2017 Jan;13(1):25-36.

Peters, T.J., Buckley, M.J., Statham, A.L. et al. De novo identification of differentially methylated regions in the human genome. *Epigenetics & chromatin* 8, 6 (2015).

Peters TJ, Buckley MJ, Chen Y, Smyth GK, Goodnow CC, Clark SJ. Calling differentially methylated regions from whole genome bisulphite sequencing with DMRcate. *Nucleic Acids Res.* 2021 Nov 8;49(19):e109.

Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Djik S, Muhlhausler B, Stirzaker C, Clark SJ. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* 2016 Oct 7;17(1):208.

Polanczyk MJ, Carson BD, Subramanian S, Afentoulis M, Vandenbark AA, Ziegler SF, et al. Cutting edge: estrogen drives expansion of the CD4⁺CD25⁺ regulatory T cell compartment. *J Immunol.* (2004) 173:2227– 30.

Portaccio E, Bellinvia A, Fonderico M, Pastò L, Razzolini L, Totaro R, Spitaleri D, Lugaresi A, Cocco E, Onofrij M, Di Palma F, Patti F, Maimone D, Valentino P, Confalonieri P, Protti A, Sola P, Lus G, Maniscalco GT, Brescia Morra V, Salemi G, Granella F, Pesci I, Bergamaschi R, Aguglia U, Vianello M, Simone M, Lepore V, Iaffaldano P, Filippi M, Trojano M, Amato MP. Progression is independent of relapse activity in early multiple sclerosis: a real-life cohort study. *Brain*. 2022 Aug 27;145(8):2796-2805.

Prosperini L, Mancinelli CR, Solaro CM, Nociti V, Haggiag S, Cordioli C, De Giglio L, De Rossi N, Galgani S, Rasia S, Ruggieri S, Tortorella C, Capra R, Mirabella M, Gasperini C. Induction Versus Escalation in Multiple Sclerosis: A 10-Year Real World Study. *Neurotherapeutics*. 2020 Jul;17(3):994-1004.

Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, Doig A, Guilliams T, Latimer J, McNamee C, Norris A, Sanseau P, Cavalla D, Pirmohamed M. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov*. 2019 Jan;18(1):41-58.

Ramanujam R, Hedström AK, Manouchehrinia A, Alfredsson L, Olsson T, Bottai M, Hillert J. Effect of Smoking Cessation on Multiple Sclerosis Prognosis. *JAMA Neurol*. 2015 Oct;72(10):1117-23.

Revez JA, Lin T, Qiao Z, Xue A, Holtz Y, Zhu Z, Zeng J, Wang H, Sidorenko J, Kemper KE, Vinkhuyzen AAE, Frater J, Eyles D, Burne THJ, Mitchell B, Martin NG, Zhu G, Visscher PM, Yang J, Wray NR, McGrath JJ. Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. *Nat Commun*. 2020 Apr 2;11(1):1647.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015). “limma powers differential expression analyses for RNA-sequencing and microarray studies.” *Nucleic Acids Research*, 43(7), e47.

Rocca MA, Battaglini M, Benedict RH, De Stefano N, Geurts JJ, Henry RG, Horsfield MA, Jenkinson M, Pagani E, Filippi M. Brain MRI atrophy quantification in MS: From methods to clinical application. *Neurology*. 2017 Jan 24;88(4):403-413.

Ross JP, van Dijk S, Phang M, Skilton MR, Molloy PL, Oytam Y. Batch-effect detection, correction and characterisation in Illumina HumanMethylation450 and MethylationEPIC BeadChip array data. *Clin Epigenetics*. 2022 Apr 29;14(1):58.

Rotstein DL, Healy BC, Malik MT, Carruthers RL, Musallam AJ, Kivisakk P, Weiner HL, Glanz B, Chitnis T. Effect of vitamin D on MS activity by disease-modifying therapy class. *Neurol Neuroimmunol Neuroinflamm*. 2015 Oct 29;2(6):e167.

Roxburgh RH, Seaman SR, Masterman T, Hensiek AE, Sawcer SJ, Vukusic S, Achiti I, Confavreux C, Coustans M, le Page E, Edan G, McDonnell GV, Hawkins S, Trojano M, Liguori M, Cocco E, Marrosu MG, Tesser F, Leone MA, Weber A, Zipp F, Milterski B, Eppelen JT, Oturai A, Sørensen PS, Celius EG, Lara NT, Montalban X, Villoslada P, Silva AM, Marta M, Leite I, Dubois B, Rubio J, Butzkueven H, Kilpatrick T, Mycko MP, Selmaj KW, Rio ME, Sá M, Salemi G, Savettieri G, Hillert J, Compston DA. Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity. *Neurology*. 2005 Apr 12;64(7):1144-51.

Saltytė Benth J, Myhr KM, Løken-Amsrud KI, Beiske AG, Bjerve KS, Hovdal H, Midgard R, Holmøy T. Modelling and prediction of 25-hydroxyvitamin D levels in Norwegian relapsing-remitting multiple sclerosis patients. *Neuroepidemiology*. 2012;39(2):84-93.

Salzer J, Hallmans G, Nystrom M, Stenlund H, Wadell G & Sundstrom P (2012) Vitamin D as a protective factor in multiple sclerosis. *Neurology* 79: 2140–2145

Sargsyan SA, Shearer AJ, Ritchie AM, Burgoon MP, Anderson S, Hemmer B, et al. Absence of Epstein-Barr virus in the brain and CSF of patients with multiple sclerosis. *Neurology* (2010) 74:1127–35.

Scalfari A, Neuhaus A, Daumer M, Muraro PA, Ebers GC. Onset of secondary progressive phase and long-term evolution of multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2014; 85: 67–75.

Schirmer L, Velmeshev D, Holmqvist S, Kaufmann M, Werneburg S, Jung D, Vistnes S, Stockley JH, Young A, Steindel M, Tung B, Goyal N, Bhaduri A, Mayer S, Engler JB, Bayraktar OA, Franklin RJM, Haeussler M, Reynolds R, Schafer DP, Friese MA, Shiow LR, Kriegstein AR, Rowitch DH. Neuronal vulnerability and multilineage diversity in multiple sclerosis. *Nature*. 2019 Sep;573(7772):75-82.

Sena A, Couderc R, Vasconcelos JC, Ferret-Sena V, Pedrosa R. Oral contraceptive use and clinical outcomes in patients with multiple sclerosis. *J Neurol Sci*. 2012 Jun 15;317(1-2):47-51.

Sepúlveda M, Ros C, Martínez-Lapiscina EH, Solà-Valls N, Hervàs M, Llufríu S, La Puma D, Casals E, Blanco Y, Villoslada P, Graus F, Castelo-Branco C, Saiz A. Pituitary-ovary axis and ovarian reserve in fertile women with multiple sclerosis: A pilot study. *Mult Scler*. 2016 Apr;22(4):564-8.

Sinnecker T, Clarke MA, Meier D, Enzinger C, Calabrese M, De Stefano N, Pitiot A, Giorgio A, Schoonheim MM, Paul F, Pawlak MA, Schmidt R, Kappos L, Montalban X, Rovira À, Evangelou N, Wuerfel J; MAGNIMS Study Group. Evaluation of the Central Vein Sign as a Diagnostic Imaging Biomarker in Multiple Sclerosis. *JAMA Neurol*. 2019 Dec 1;76(12):1446-1456.

Smolders J, Torkildsen Ø, Camu W, Holmøy T. An Update on Vitamin D and Disease Activity in Multiple Sclerosis. *CNS Drugs*. 2019 Dec;33(12):1187-1199. University of California, San Francisco MS-EPIC Team; Cree BA, Gourraud PA, Oksenberg JR, Bevan C, Crabtree-Hartman E, Gelfand JM, Goodin DS, Graves J, Green AJ, Mowry E, Okuda DT, Pelletier D, von Büdingen HC, Zamvil SS, Agrawal A, Caillier S, Ciocca C, Gomez R, Kanner R, Lincoln R, Lizee A, Qualley P, Santaniello A, Suleiman L, Bucci M, Panara V, Papinutto N, Stern WA, Zhu AH, Cutter GR, Baranzini S, Henry RG, Hauser SL. Long-term evolution of multiple sclerosis disability in the treatment era. *Ann Neurol*. 2016 Oct;80(4):499-510.

Sommer K, Guo B, Pomerantz JL, Bandaranayake AD, Moreno-García ME, Ovechkina YL, Rawlings DJ. Phosphorylation of the CARMA1 linker controls NF- κ B activation. *Immunity*. 2005 Dec;23(6):561-74.

Stephenson E, Nathoo N, Mahjoub Y, Dunn JF, Yong VW. Iron in multiple sclerosis: roles in neurodegeneration and repair. *Nat Rev Neurol*. 2014 Aug;10(8):459-68.

Stewart N, Simpson S Jr, van der Mei I, Ponsonby AL, Blizzard L, Dwyer T, Pittas F, Eyles D, Ko P, Taylor BV. Interferon- β and serum 25-hydroxyvitamin D interact to modulate relapse risk in MS. *Neurology*. 2012 Jul 17;79(3):254-60.

Sun H, Wilman AH. Background field removal using spherical mean value filtering and Tikhonov regularization. *Magn Reson Med*. 2014;71(3):1151-7.

Sun H, Wilman AH. Quantitative susceptibility mapping using single-shot echo-planar imaging. *Magn Reson Med*. 2015 May;73(5):1932-8.

Sun Z, Chai HS, Wu Y, White WM, Donkena KV, Klein CJ, Garovic VD, Therneau TM, Kocher JP. Batch effect correction for genome-wide methylation data with Illumina Infinium platform. *BMC Med Genomics*. 2011 Dec 16;4:84.

Sureshchandra S, Raus A, Jankeel A, Ligh BJK, Walter NAR, Newman N, Grant KA, Messaoudi I. Dose-dependent effects of chronic alcohol drinking on peripheral immune responses. *Sci. Rep.* 2019, 9, 7847.

Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, Correale J, Fazekas F, Filippi M, Freedman MS, Fujihara K, Galetta SL, Hartung HP, Kappos L, Lublin FD, Marrie RA, Miller AE, Miller DH, Montalban X, Mowry EM, Sorensen PS, Tintoré M, Traboulsee AL, Trojano M, Uitdehaag BMJ, Vukusic S, Waubant E, Weinshenker BG, Reingold SC, Cohen JA. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol*. 2018 Feb;17(2):162-173.

Thompson EE, Nicodemus-Johnson J, Kim KW, Gern JE, Jackson DJ, Lemanske RF, et al. Global DNA methylation changes spanning puberty are near predicted estrogen-responsive genes and enriched for genes involved in endocrine and immune processes. *Clin Epigenetics*. (2018) 10:62. doi: 10.1186/s13148-018-0491-2.

Thöne J, Kollar S, Noursome D, Ellrichmann G, Kleiter I, Gold R, Hellwig K. Serum anti-Müllerian hormone levels in reproductive-age women with relapsing-remitting multiple sclerosis. *Mult Scler*. 2015 Jan;21(1):41-7.

Tintore M, Rovira À, Ríó J, Otero-Romero S, Arrambide G, Tur C, Comabella M, Nos C, Arévalo MJ, Negrotto L, Galán I, Vidal-Jordana A, Castelló J, Palavra F, Simon E, Mitjana R, Auger C, Sastre-Garriga J, Montalban X. Defining high, medium and low impact prognostic factors for developing multiple sclerosis. *Brain*. 2015 Jul;138(Pt 7):1863-74.

Ucciferri CC, Dunn SE. Effect of puberty on the immune system: Relevance to multiple sclerosis. *Front Pediatr*. 2022 Dec 2;10:1059083.

Valverde S, Cabezas M, Roura E, Gonzalez-Villa S, Pareto D, Vilanova JC, et al. Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. *Neuroimage*. 2017;155:159-68.

Van Hijfte L, Loret G, Bachmann H, Reynders T, Breuls M, Deschepper E, Kuhle J, Willekens B, Laureys G. Lifestyle factors in multiple sclerosis disability progression and

silent brain damage: A cross-sectional study. *Mult Scler Relat Disord*. 2022 Sep;65:104016.

Vandebergh M, Andlauer TFM, Zhou Y, Mallants K, Held F, Aly L, Taylor BV, Hemmer B, Dubois B, Goris A. Genetic Variation in WNT9B Increases Relapse Hazard in Multiple Sclerosis. *Ann Neurol*. 2021 May;89(5):884-894.

Vandebergh M, Dubois B, Goris A. Effects of Vitamin D and Body Mass Index on Disease Risk and Relapse Hazard in Multiple Sclerosis: A Mendelian Randomization Study. *Neurol Neuroimmunol Neuroinflamm*. 2022 Apr 7;9(3):e1165.

Verdin E. NAD⁺ in aging, metabolism, and neurodegeneration. *Science*. 2015 Dec 4;350(6265):1208-13.

Võsa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet*. 2021 Sep;53(9):1300-1310.

Walton C, King R, Rechtman L, Kaye W, Leray E, Marrie RA, Robertson N, La Rocca N, Uitdehaag B, van der Mei I, Wallin M, Helme A, Angood Napier C, Rijke N, Baneke P. Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. *Mult Scler*. 2020 Dec;26(14):1816-1821.

Wang C, Dehghani B, Li Y, Kaler LJ, Vandebark AA, Offner H. Oestrogen modulates experimental autoimmune encephalomyelitis and interleukin-17 production via programmed death 1. *Immunology*. (2009) 126:329–35.

Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, et al. Distribution, silencing potential, and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet*. 2007;39:457–66.

Wesnes K, Riise T, Casetta I, Drulovic J, Granieri E, Holmøy T, Kampman MT, Landtblom AM, Lauer K, Lossius A, Magalhaes S, Pekmezovic T, Bjørnevik K, Wolfson C, Pugliatti M, Myhr KM. Body size and the risk of multiple sclerosis in Norway and Italy: the EnvIMS study. *Mult Scler*. 2015 Apr;21(4):388-95.

Westerlind Helga and Manouchehrinia Ali. Ms.sev: Package for Calculation of ARMSS, Local MSSS and Global MSSS. 2016 Dec. <https://cran.r-project.org/web/packages/ms.sev>

Williams A, Piaton G, Aigrot MS, Belhadi A, Théaudin M, Petermann F, Thomas JL, Zalc B, Lubetzki C. Semaphorin 3A and 3F: key players in myelin repair in multiple

sclerosis? *Brain*. 2007 Oct;130(Pt 10):2554-65. doi: 10.1093/brain/awm202. Epub 2007 Sep 11.

Wood CL, Lane LC, Cheetham T. Puberty: normal physiology (brief overview). *Best Pract Res Clin Endocrinol Metab*. (2019) 33:101265.

Wu Q, Wang Q, Mao G, Dowling CA, Lundy SK, Mao-Draayer Y. Dimethyl Fumarate Selectively Reduces Memory T Cells and Shifts the Balance between Th1/Th17 and Th2 in Multiple Sclerosis Patients. *J Immunol*. 2017 Apr 15;198(8):3069-3080.

Xu HY, Zhang HX, Xiao Z, Qiao J, Li R. Regulation of anti-Müllerian hormone (AMH) in males and the associations of serum AMH with the disorders of male fertility. *Asian J Androl*. 2019 Mar-Apr;21(2):109-114.

Ysraelit MC, Correale J. Impact of sex hormones on immune function and multiple sclerosis development. *Immunology*. 2019 Jan;156(1):9-22.

Zhang Z, Yao L, Yang J, Wang Z, Du G. PI3K/Akt and HIF-1 signaling pathway in hypoxia-ischemia (Review). *Mol Med Rep*. 2018 Oct;18(4):3547-3554.

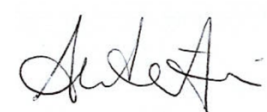
Zhao G, Liu Y, Fang J, Chen Y, Li H, Gao K. Dimethyl fumarate inhibits the expression and function of hypoxia-inducible factor-1 α (HIF-1 α). *Biochem Biophys Res Commun*. 2014 Jun 6;448(3):303-7.

Zheleznyakova GY, Piket E, Marabita F, Pahlevan Kakhki M, Ewing E, Ruhrmann S, Needhamsen M, Jagodic M, Kular L. Epigenetic research in multiple sclerosis: progress, challenges, and opportunities. *Physiol Genomics*. 2017 Sep 1;49(9):447-461.

Zheng SC, Breeze CE, Beck S, Teschendorff AE. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat Methods*. 2018 Dec;15(12):1059-1066.

Zhou DD, Luo M, Huang SY, Saimaiti A, Shang A, Gan RY, Li HB. Effects and Mechanisms of Resveratrol on Aging and Age-Related Diseases. *Oxid Med Cell Longev*. 2021 Jul 11;2021:9932218.

Zou Y, Zhang WF, Liu HY, Li X, Zhang X, Ma XF, Sun Y, Jiang SY, Ma QH, Xu DE. Structure and function of the contactin-associated protein family in myelinated axons and their relationship with nerve diseases. *Neural Regen Res*. 2017 Sep;12(9):1551-1558.



APPENDIX

Supplementary Tables

Table S3.1 & Table S3.1

<i>CHR</i>	<i>SNP</i>	<i>BP</i>	<i>AI</i>	<i>NMISS</i>	<i>OR</i>	<i>P</i>	<i>Gene</i>
15	rs11633419	23779479	T	1152	0.639	7.80E-07	MKRN3
15	rs4778558	23782995	T	1147	0.640	7.88E-07	MKRN3
15	rs10431854	23780472	C	1152	0.639	7.98E-07	MKRN3
15	rs11638266	23781646	G	1152	0.639	7.98E-07	MKRN3
15	rs11858970	23777111	G	1151	0.640	8.88E-07	MKRN3
15	rs8036462	23777880	G	1153	0.641	9.17E-07	MKRN3
7	rs2158725	83297241	C	1151	0.533	1.46E-06	SEMA3E
7	rs9649281	105985813	T	1149	0.626	2.20E-06	NAMPT
7	rs9649282	105986115	A	1149	0.626	2.20E-06	NAMPT
7	rs6975444	105973847	T	1150	0.632	3.34E-06	NAMPT
7	rs6975772	105974057	T	1150	0.632	3.34E-06	NAMPT
16	rs1549660	76367186	T	1145	0.652	3.47E-06	CNTNAP4
3	rs17785714	167897639	T	1152	0.403	3.84E-06	GOLIM4
3	rs75910889	167898645	A	1152	0.403	3.84E-06	GOLIM4
10	rs920259	60396065	A	1154	0.625	4.24E-06	BICC1
16	rs9934361	76354742	C	1151	0.66	4.68E-06	CNTNAP4
5	rs10462567	78630605	A	1150	1.55	4.72E-06	JMY
3	rs41464047	167896971	T	1152	0.41	5.48E-06	GOLIM4
3	rs57169293	167897194	G	1152	0.41	5.48E-06	GOLIM4
3	rs9968241	167897462	A	1152	0.41	5.48E-06	GOLIM4

Table S3.1. Results from the GWAS in the OSR cohort. The 20 top-associated SNPs are reported. *NMISS* = Number of non-missing genotypes. *OR* = odds ratio for the EDA status. *Gene* = University of California Santa Cruz (UCSC) gene symbol for the gene with the highest score following the OpenTarget Genetics pipeline.

<i>CHR</i>	<i>SNP</i>	<i>BP</i>	<i>AI</i>	<i>NMISS</i>	<i>OR</i>	<i>P</i>	<i>Gene</i>
15	rs11632455	60126990	T	250	0.32	4.44E-06	FOXB1
4	rs2287145	62944272	A	253	0.17	4.85E-06	ADGRL3
13	rs8002310	75182628	A	254	0.27	5.41E-06	KLF12
5	rs4836114	124082582	C	250	3.13	6.00E-06	ZNF608
13	rs7331788	75165636	C	254	0.28	6.63E-06	KLF12
13	rs9573423	75187987	C	254	0.28	9.49E-06	KLF12
13	rs9573418	75171074	G	254	0.30	1.37E-05	KLF12
13	rs9565107	75175663	C	254	0.30	1.37E-05	KLF12
1	rs61830764	212289976	A	248	0.39	1.43E-05	DTL
3	rs35844296	3743992	A	249	0.27	1.47E-05	LRRN1
17	rs35814610	13041714	C	251	0.36	1.66E-05	ELAC2
17	rs8077659	13042271	A	251	0.36	1.66E-05	ELAC2
17	rs8077563	13042365	T	251	0.36	1.66E-05	ELAC2
17	rs8077971	13042481	A	251	0.36	1.66E-05	ELAC2
19	rs746641	33369655	C	254	0.35	1.87E-05	FAAP24
8	rs28534381	40324892	A	249	0.35	1.90E-05	TCIM
13	rs7338814	75174084	C	254	0.31	1.91E-05	KLF12
10	rs2484191	27694764	T	254	0.17	2.05E-05	MASTL
2	rs111631454	210092724	C	253	0.22	2.05E-05	MAP2
17	rs764191	13039792	T	251	0.36	2.08E-05	ELAC2

Table S3.2. Results from the GWAS in the CHUT cohort. The 20 top-associated SNPs are reported. NMISS = Number of non-missing genotypes. OR = odds ratio for the EDA status. Gene = University of California Santa Cruz (UCSC) gene symbol for the gene with the highest score following the OpenTarget Genetics pipeline. For rs2287145 the nearest gene was selected, as the variant was missing in the OpenTarget repository

Table S7.1 & Table S7.2

CHR	START	END	GENE	START-2kb	START+2kb
<i>1</i>	11994723	12035599	PLOD1	11992723	12037599
<i>1</i>	17312452	17338423	ATP13A2	17310452	17340423
<i>1</i>	22963117	22966175	C1QA	22961117	22968175
<i>1</i>	43212005	43232755	P3H1	43210005	43234755
<i>1</i>	44440601	44443972	ATP6V0B	44438601	44445972
<i>1</i>	44462154	44483012	SLC6A9	44460154	44485012
<i>1</i>	47264669	47285021	CYP4B1	47262669	47287021
<i>1</i>	47308766	47366147	CYP4Z2P	47306766	47368147
<i>1</i>	47394845	47407156	CYP4A11	47392845	47409156
<i>1</i>	47489239	47516423	CYP4X1	47487239	47518423
<i>1</i>	47533159	47583992	CYP4Z1	47531159	47585992
<i>1</i>	47603106	47614526	CYP4A22	47601106	47616526
<i>1</i>	60358979	60392423	CYP2J2	60356979	60394423
<i>1</i>	94352589	94375012	GCLM	94350589	94377012
<i>1</i>	109648572	109656479	C1orf194	109646572	109658479
<i>1</i>	145413190	145417545	HJV	145411190	145419545
<i>1</i>	155259083	155271225	PKLR	155257083	155273225
<i>1</i>	198492351	198510075	ATP6V1G3	198490351	198512075
<i>1</i>	213031596	213072705	FLVCR1	213029596	213074705
<i>1</i>	220087605	220101993	SLC30A10	220085605	220103993
<i>1</i>	228353428	228369958	IBA57	228351428	228371958
<i>1</i>	231499496	231560790	EGLN1	231497496	231562790
<i>1</i>	235530727	235612280	TBCE	235528727	235614280
<i>1</i>	242158791	242162385	MAP1LC3C	242156791	242164385
<i>2</i>	3501689	3523350	ADI1	3499689	3525350
<i>2</i>	10861774	10925236	ATP6V1C2	10859774	10927236
<i>2</i>	31557187	31637611	XDH	31555187	31639611
<i>2</i>	38294745	38303323	CYP1B1	38292745	38305323
<i>2</i>	42994228	43019751	HAAO	42992228	43021751
<i>2</i>	46524540	46613842	EPAS1	46522540	46615842
<i>2</i>	46755024	46769141	ATP6V1E2	46753024	46771141
<i>2</i>	47168312	47303275	TTC7A	47166312	47305275

2	69623244	69664760	NFU1	69621244	69666760
2	70314584	70316334	PCBP1	70312584	70318334
2	71162997	71192561	ATP6V1B1	71160997	71194561
2	72356366	72374991	CYP26B1	72354366	72376991
2	86668583	86719839	KDM3A	86666583	86721839
2	96931883	96939900	CIAO1	96929883	96941900
2	119981383	120023227	STEAP3	119979383	120025227
2	127941411	127963343	CYP27C1	127939411	127965343
2	131095815	131099922	CCDC115	131093815	131101922
2	172378865	172414643	CYBRD1	172376865	172416643
2	190425315	190445537	SLC40A1	190423315	190447537
2	200793633	200820459	TYW5	200791633	200822459
2	201450730	201536217	AOX1	201448730	201538217
2	204103163	204170563	CYP20A1	204101163	204172563
2	219246751	219261617	SLC11A1	219244751	219263617
2	219524378	219528166	BCS1L	219522378	219530166
2	219646471	219680016	CYP27A1	219644471	219682016
2	220074487	220085174	RP11-803J6.1, ABCB6	220072487	220087174
2	223725731	223808119	ACSL3	223723731	223810119
2	239072632	239077515	ERFE	239070632	239079515
3	3168599	3190706	TRNT1	3166599	3192706
3	11314009	11599139	ATG7	11312009	11601139
3	42913683	42917633	CYP8B1	42911683	42919633
3	46477495	46506598	LTF	46475495	46508598
3	48601505	48632593	COL7A1	48599505	48634593
3	49027340	49044581	P4HTM	49025340	49046581
3	50355220	50360281	HYAL2	50353220	50362281
3	71003864	71180092	FOXP1	71001864	71182092
3	113465865	113530905	ATP6V1A	113463865	113532905
3	133464799	133497850	TF	133462799	133499850
3	145787227	145879282	PLOD2	145785227	145881282
3	148890289	148939832	CP	148888289	148941832
3	184428154	184429836	MAGEF1	184426154	184431836
3	186383797	186396023	HRG	186381797	186398023

3	189674516	189838908	P3H2	189672516	189840908
3	195776154	195809032	TFRC	195774154	195811032
3	196728611	196756687	MELTF	196726611	196758687
4	15606006	15657035	FBXL5	15604006	15659035
4	76781025	76823681	PPEF2	76779025	76825681
4	83550689	83720010	SCD5	83548689	83722010
4	89011415	89080011	ABCG2	89009415	89082011
4	90645249	90759447	SNCA	90643249	90761447
4	103182820	103266655	SLC39A8	103180820	103268655
4	103998781	104021024	BDH2	103996781	104023024
4	106067841	106200960	TET2	106065841	106202960
4	108852716	108874613	CYP2U1	108850716	108876613
4	129190391	129209984	PGRMC2	129188391	129211984
4	139085247	139163503	SLC7A11	139083247	139165503
4	146019155	146050676	ABCE1	146017155	146052676
4	166248817	166264314	MSMO1	166246817	166266314
4	185676748	185747215	ACSL1	185674748	185749215
4	187112673	187134617	CYP4V2	187110673	187136617
5	1392904	1445543	SLC6A3	1390904	1447543
5	68462836	68474070	CCNB1	68460836	68476070
5	94799598	94890709	TTC37	94797598	94892709
5	115140429	115152405	CDO1	115138429	115154405
5	121187649	121188523	FTMT	121185649	121190523
5	131285666	131347355	ACSL6	131283666	131349355
5	131584600	131631008	P4HA2	131582600	131633008
5	133492081	133512724	SKP1	133490081	133514724
5	137890570	137911318	HSPA9	137888570	137913318
5	141488323	141534008	NDFIP1	141486323	141536008
5	154198051	154230213	FAXDC2	154196051	154232213
5	172410762	172461900	ATP6V0E1	172408762	172463900
6	5186833	5261172	LYRM4	5184833	5263172
6	7727010	7881961	BMP6	7725010	7883961
6	10396915	10415470	TFAP2A	10394915	10417470
6	26087508	26095469	HFE	26085508	26097469
6	31512227	31514625	ATP6V1G2	31510227	31516625

6	31926580	31937532	SKIV2L	31924580	31939532
6	31973358	31976712	CYP21A1P, CYP21A	31971358	31978712
6	32006092	32009447	CYP21A2	32004092	32011447
6	38136226	38607924	BTBD9	38134226	38609924
6	46517444	46620523	CYP39A1	46515444	46622523
6	53362139	53409927	GCLC	53360139	53411927
6	106632351	106773695	ATG5	106630351	106775695
7	1022834	1029276	CYP2W1	1020834	1031276
7	6061877	6098860	EIF2AK1	6059877	6100860
7	15239942	15601640	AGMO	15237942	15603640
7	39606002	39612480	YAE1	39604002	39614480
7	87834431	87849399	SRI	87832431	87851399
7	87905743	87936228	STEAP4	87903743	87938228
7	89783688	89794141	STEAP1	89781688	89796141
7	89840999	89866992	STEAP2	89838999	89868992
7	91741462	91763840	CYP51A1	91739462	91765840
7	97736196	97838944	LMTK2	97734196	97840944
7	99282301	99332819	CYP3A7- CYP3A51P, CYP3A7	99280301	99334819
7	99354582	99381811	CYP3A4	99352582	99383811
7	99425635	99464173	CYP3A43	99423635	99466173
7	100218038	100239201	TFR2	100216038	100241201
7	100849257	100861011	PLOD3	100847257	100863011
7	128502856	128505903	ATP6V1F	128500856	128507903
7	138391038	138458782	ATP6V0A4	138389038	138460782
7	139528951	139720125	TBXAS1	139526951	139722125
7	139784545	139876741	KDM7A	139782545	139878741
7	148395932	148498202	CUL1	148393932	148500202
7	149570056	149577787	ATP6V0E2	149568056	149579787
7	150745378	150749843	ASIC3	150743378	150751843
8	20054703	20079207	ATP6V1B2	20052703	20081207
8	22225049	22280249	SLC39A14	22223049	22282249
8	23386362	23430063	SLC25A37	23384362	23432063

8	27727398	27850369	SCARA5	27725398	27852369
8	42249278	42263455	VDAC3	42247278	42265455
8	54628102	54755871	ATP6V1H	54626102	54757871
8	59402736	59412720	CYP7A1	59400736	59414720
8	65508528	65711348	CYP7B1	65506528	65713348
8	87111138	87166454	ATP6V0D2	87109138	87168454
8	104033247	104085285	ATP6V1C1	104031247	104087285
8	128748314	128753680	MYC	128746314	128755680
8	143953772	143961236	CYP11B1	143951772	143963236
8	143991974	143999259	CYP11B2	143989974	144001259
9	32384600	32450832	ACO1	32382600	32452832
9	71650478	71693993	FXN	71648478	71695993
9	79792360	80032399	VPS13A	79790360	80034399
9	88879462	88897490	ISCA1	88877462	88899490
9	96338908	96441869	PHF2	96336908	96443869
9	116148591	116163618	ALAD	116146591	116165618
9	117349993	117361152	ATP6V1G1	117347993	117363152
9	130911731	130915734	LCN2	130909731	130917734
9	139257440	139268133	CARD9	139255440	139270133
9	140100118	140113813	NDOR1	140098118	140115813
10	13319795	13342130	PHYH	13317795	13344130
10	45869623	45941567	ALOX5	45867623	45943567
10	48413091	48416853	GDF2	48411091	48418853
10	51565107	51590734	NCOA4	51563107	51592734
10	70320116	70454239	TET1	70318116	70456239
10	74766979	74856732	P4HA1	74764979	74858732
10	76969911	76991207	VDAC2	76967911	76993207
10	90965693	90967071	CH25H	90963693	90969071
10	94821020	94828454	CYP26C1	94819020	94830454
10	94833646	94837641	CYP26A1	94831646	94839641
10	96522462	96612671	CYP2C19	96520462	96614671
10	96698414	96749148	CYP2C9	96696414	96751148
10	96796528	96829254	CYP2C8	96794528	96831254
10	99218080	99258366	MMS19	99216080	99260366
10	101370274	101380221	SLC25A28	101368274	101382221

10	102106771	102124588	SCD	102104771	102126588
10	102295640	102313681	HIF1AN	102293640	102315681
10	104590287	104597290	CYP17A1	104588287	104599290
10	114135955	114188138	ACSL5	114133955	114190138
10	131934638	131977932	GLRX3	131932638	131979932
10	135340866	135352620	CYP2E1	135338866	135354620
11	568088	568198	MIR210	566088	570198
11	2185158	2193035	TH	2183158	2195035
11	5246695	5248301	HBB	5244695	5250301
11	6452267	6462254	HPX	6450267	6464254
11	14899555	14913751	CYP2R1	14897555	14915751
11	18042083	18062335	TPH1	18040083	18064335
11	27062508	27149354	BBOX1	27060508	27151354
11	31391376	31454382	DNAJC24	31389376	31456382
11	34937676	35017675	PDHXPDX1	34935676	35019675
11	43902356	43941825	ALKBH3	43900356	43943825
11	46698624	46722215	ARHGAP1	46696624	46724215
11	61731756	61735132	FTH1	61729756	61737132
11	62623483	62656355	SLC3A2	62621483	62658355
11	67806461	67818366	TCIRG1	67804461	67820366
11	69455872	69469242	CCND1	69453872	69471242
11	69480331	69490165	LTO1	69478331	69492165
11	73977701	74022699	P4HA3	73975701	74024699
11	85668213	85780139	PICALM	85666213	85782139
11	93754377	93847374	HEPHL1	93752377	93849374
11	107373452	107436461	ALKBH8	107371452	107438461
11	110300660	110335608	FDX1	110298660	110337608
11	111895537	111935002	DLAT	111893537	111937002
11	113280316	113346001	DRD2	113278316	113348001
11	119531702	119599435	NECTIN1	119529702	119601435
11	121163387	121184119	SC5D	121161387	121186119
12	7085346	7125842	LPCAT3	7083346	7127842
12	27849427	27850566	REP15	27847427	27852566
12	46576840	46663208	SLC38A1, SAT1	46574840	46665208
12	46751970	46766645	SLC38A2, SAT2	46749970	46768645

12	48166966	48176536	SLC48A1	48164966	48178536
12	51379774	51422058	SLC11A2	51377774	51424058
12	53845885	53874946	PCBP2	53843885	53876946
12	58156116	58160976	CYP27B1	58154116	58162976
12	67663060	67708388	CAND1	67661060	67710388
12	68548549	68553521	IFNG	68546549	68555521
12	72332625	72426221	TPH2	72330625	72428221
12	103232103	103311381	PAH	103230103	103313381
12	108956293	108963160	ISCU	108954293	108965160
12	109525992	109531293	ALKBH2	109523992	109533293
12	116997185	117014425	MAP1LC3B2	116995185	117016425
12	123459353	123464588	OGFOD2	123457353	123466588
12	124196864	124246301	ATP6V0A2	124194864	124248301
13	28494167	28500451	PDX1	28492167	28502451
14	23815526	23821660	SLC22A17	23813526	23823660
14	32030590	32330429	NUBPL	32028590	32332429
14	34393420	34420284	EGLN3	34391420	34422284
14	62162118	62214977	HIF1A	62160118	62216977
14	67804580	67826720	ATP6V1D	67802580	67828720
14	74960422	74962271	ISCA2	74958422	74964271
14	76044939	76114512	FLVCR2	76042939	76116512
14	78138748	78174356	ALKBH1	78136748	78176356
14	96001322	96011055	GLRX5	95999322	96013055
14	100150754	100193638	CYP46A1	100148754	100195638
15	43489425	43513323	EPB42	43487425	43515323
15	45003684	45010357	B2M	45001684	45012357
15	51500253	51630795	CYP19A1	51498253	51632795
15	64364760	64386207	CIAO2A	64362760	64388207
15	69307033	69349501	NOX5	69305033	69351501
15	69706626	69740764	KIF23	69704626	69742764
15	73344824	73597547	NEO1	73342824	73599547
15	74630102	74660081	CYP11A1	74628102	74662081
15	75011882	75017877	CYP1A1	75009882	75019877
15	75041183	75048941	CYP1A2	75039183	75050941
15	78730517	78793798	IREB2	78728517	78795798

16	202853	204504	HBZ	200853	206504
16	222845	223709	HBA2	220845	225709
16	226678	227520	HBA1	224678	229520
16	230332	231178	HBQ1	228332	233178
16	779768	790997	CIAO3	777768	792997
16	1832932	1839192	NUBP2	1830932	1841192
16	2563726	2570224	ATP6V0C	2561726	2572224
16	4526340	4560348	HMOX2	4524340	4562348
16	10837697	10863208	NUBP1	10835697	10865208
16	29464913	29466285	BOLA2	29462913	29468285
16	30204255	30205627	BOLA2B	30202255	30207627
16	53737874	54148379	FTO	53735874	54150379
16	56485423	56511407	OGFOD1	56483423	56513407
16	57462086	57481369	CIAPIN1	57460086	57483369
16	66965957	66968326	CIAO2B	66963957	66970326
16	67471916	67515089	ATP6V0D1	67469916	67517089
16	74746855	74808729	FA2H	74744855	74810729
16	87425800	87438380	MAP1LC3B	87423800	87440380
17	4534213	4544971	ALOX15	4532213	4546971
17	6899383	6914055	ALOX12	6897383	6916055
17	7529555	7531194	SAT2	7527555	7533194
17	7571719	7590868	TP53	7569719	7592868
17	7942357	7952451	ALOX15B	7940357	7954451
17	7975953	7991021	ALOX12B	7973953	7993021
17	7999217	8022234	ALOXE3	7997217	8024234
17	26684686	26689089	TMEM199	26682686	26691089
17	26721660	26733230	SLC46A1	26719660	26735230
17	36886509	36891858	CISD3	36884509	36893858
17	40610861	40674597	ATP6V0A1	40608861	40676597
17	40962149	40976310	BECN1	40960149	40978310
17	57697049	57774317	CLTC	57695049	57776317
17	74523429	74533987	CYGB	74521429	74535987
17	74708913	74722881	JMJD6	74706913	74724881
17	80347085	80376513	OGFOD3	80345085	80378513
18	48556582	48611411	SMAD4	48554582	48613411

18	55212072	55253969	FECH	55210072	55255969
18	60790578	60986613	BCL2	60788578	60988613
19	1104648	1106787	GPX4	1102648	1108787
19	3490818	3500621	DOHH	3488818	3502621
19	7587495	7598895	MCOLN1	7585495	7600895
19	8455204	8469317	RAB11B	8453204	8471317
19	10828728	10942586	DNM2	10826728	10944586
19	11685474	11689801	ACP5	11683474	11691801
19	13049413	13055304	CALR	13047413	13057304
19	15619335	15663128	CYP4F22	15617335	15665128
19	15726028	15740447	CYP4F8	15724028	15742447
19	15751706	15771570	CYP4F3	15749706	15773570
19	15783827	15807984	CYP4F12	15781827	15809984
19	15988833	16008884	CYP4F2	15986833	16010884
19	16023179	16045676	CYP4F11	16021179	16047676
19	35773409	35776045	HAMP	35771409	35778045
19	41305333	41314346	EGLN2	41303333	41316346
19	41349442	41356352	CYP2A6	41347442	41358352
19	41381343	41388657	CYP2A7	41379343	41390657
19	41396730	41406413	CYP2A6	41394730	41408413
19	41497203	41524301	CYP2B6	41495203	41526301
19	41594355	41602100	CYP2A13	41592355	41604100
19	41620352	41634281	CYP2F1	41618352	41636281
19	41699114	41713444	CYP2S1	41697114	41715444
19	44010870	44031396	ETHE1	44008870	44033396
19	49468565	49470136	FTL	49466565	49472136
19	54926604	54947899	TTYH1	54924604	54949899
20	3869741	3904502	PANK2	3867741	3906502
20	4666796	4682234	PRNP	4664796	4684234
20	6748744	6760910	BMP2	6746744	6762910
20	33146500	33148149	MAP1LC3A	33144500	33150149
20	33516235	33543601	GSS	33514235	33545601
20	34256609	34287287	NFS1	34254609	34289287
20	48120410	48184707	PTGIS	48118410	48186707
20	52769987	52790516	CYP24A1	52767987	52792516

21	33031934	33041243	SOD1	33029934	33043243
21	43782390	43786644	TFF1	43780390	43788644
22	18074902	18111588	ATP6V1E1	18072902	18113588
22	29138042	29153496	HSCB	29136042	29155496
22	35777059	35790207	HMOX1	35775059	35792207
22	37461478	37499693	TMPRSS6	37459478	37501693
22	38507501	38577761	PLA2G6	38505501	38579761
22	41865128	41924993	ACO2	41863128	41926993
22	42522500	42526883	CYP2D6	42520500	42528883
22	42536213	42540575	CYP2D7	42534213	42542575
22	50925212	50928750	MIOX	50923212	50930750
X	18709044	18846034	PPEF1	18707044	18848034
X	23801274	23804327	SAT1	23799274	23806327
X	31089357	31090170	FTHL17	31087357	31092170
X	37639269	37672714	CYBB	37637269	37674714
X	48932091	48937564	WDR45	48930091	48939564
X	53963112	54071569	PHF8	53961112	54073569
X	55035487	55057497	ALAS2	55033487	55059497
X	65382432	65487230	HEPH	65380432	65489230
X	74273006	74376175	ABCB7	74271006	74378175
X	77166193	77305892	ATP7A	77164193	77307892
X	108884563	108976621	ACSL4	108882563	108978621
X	135044230	135056134	MMGT1	135042230	135058134
X	153656977	153664862	ATP6AP1	153654977	153666862
X	153759605	153775233	G6PD	153757605	153777233
X	154718672	154842622	TMLHE	154716672	154844622

Table S7.1. List of the mapped genes in iron metabolism included in the genetic analysis. Reference genome: GRCh37/hg19). Gene symbols and positions in base-pairs (BP) including the 2 KB flanking region are reported.

<i>Gene set</i>	<i>Source</i>	<i>N of genes</i>
<i>GOBP_IRON_ION_TRANSPORT</i>	GO	79
<i>GOBP_CELLULAR_IRON_ION_HOMEOSTASIS</i>	GO	70
<i>GOBP_RESPONSE_TO_IRON_ION</i>	GO	28
<i>GOBP_RESPONSE_TO_IRON_II_ION</i>	GO	5
<i>GOBP_RESPONSE_TO_IRON_III_ION</i>	GO	5
<i>GOBP_IRON_SULFUR_CLUSTER_ASSEMBLY</i>	GO	24
<i>GOBP_IRON_IMPORT_INTO_CELL</i>	GO	11
<i>GOBP_IRON_ION_TRANSMEMBRANE_TRANSPORT</i>	GO	19
<i>GOBP_REGULATION_OF_IRON_ION_TRANSPORT</i>	GO	9
<i>GOBP_REGULATION_OF_IRON_ION_TRANSMEMBRANE_TRANSPORT</i>	GO	6
<i>GOBP_IRON_ION_HOMEOSTASIS</i>	GO	86
<i>GOBP_MULTICELLULAR_ORGANISMAL_IRON_ION_HOMEOSTASIS</i>	GO	7
<i>GOBP_CELLULAR_RESPONSE_TO_IRON_ION</i>	GO	9
<i>GOBP_PROTEIN_MATURATION_BY_IRON_SULFUR_CLUSTER_TRANSFER</i>	GO	16
<i>GOBP_SEQUESTERING_OF_IRON_ION</i>	GO	7
<i>GOBP_IRON_ION_IMPORT_ACROSS_PLASMA_MEMBRANE</i>	GO	6
<i>GOBP_IRON_COORDINATION_ENTITY_TRANSPORT</i>	GO	14
<i>GOMF_IRON_ION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY</i>	GO	10
<i>GOMF_IRON_ION_BINDING</i>	GO	150
<i>GOMF_FERROUS_IRON_BINDING</i>	GO	26
<i>GOMF_FERRIC_IRON_BINDING</i>	GO	10

HP_ABNORMALITY_OF_IRON_HOMEOSTASIS

HP_IRON_ACCUMULATION_IN_BRAIN

KEGG_FERROPTOSIS

HP	19
HP	9
KEGG	41

Table S7.2. List of gene sets from Gene Ontology (GO), Human Phenotype (HP) and Kyoto Encyclopedia of Genes and Genome (KEGG) included in the study. *N genes* = number of genes in the gene set.

Supplementary Figures

Figure S2.1: Overview of the work

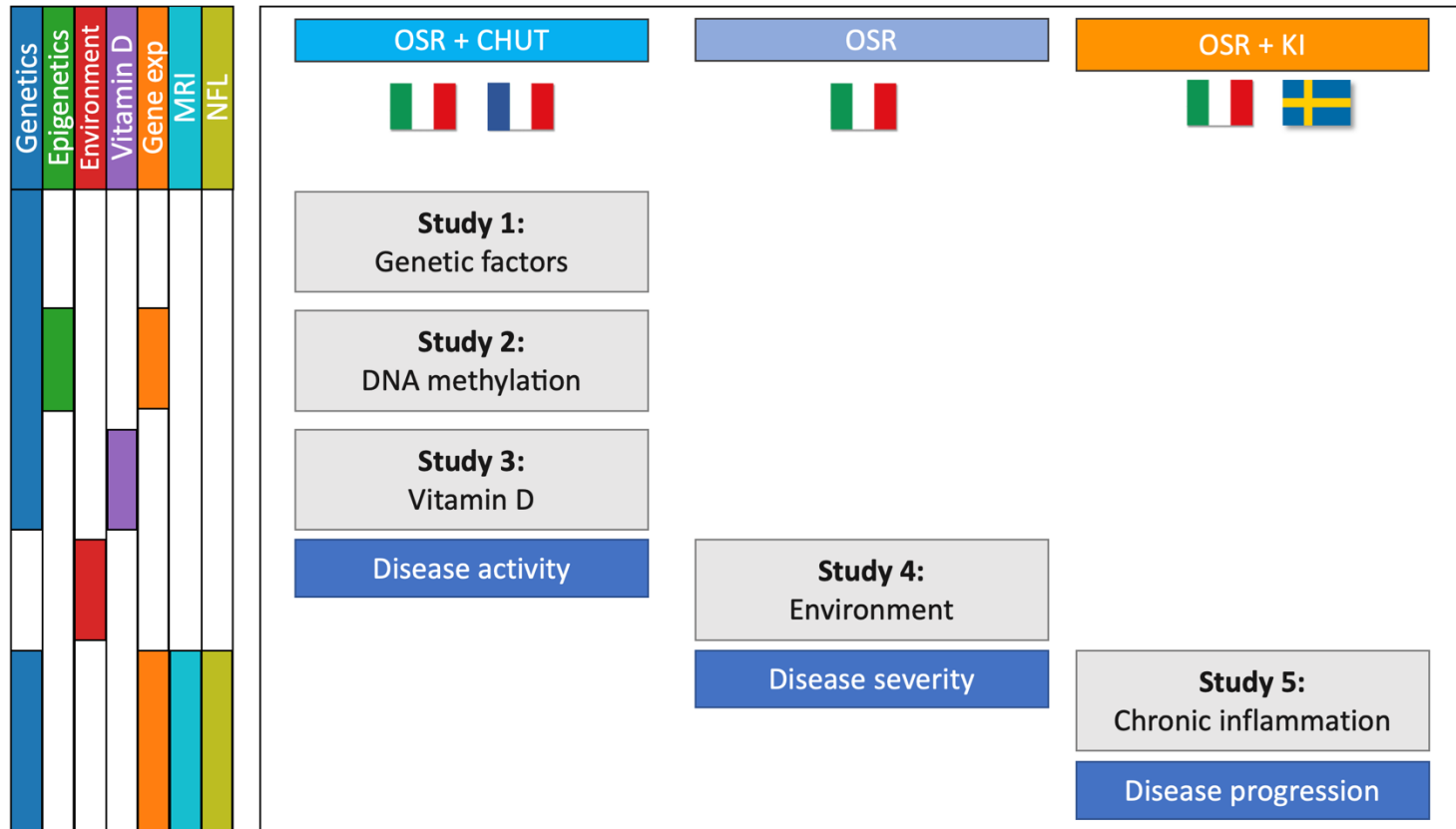



Figure S2.1. An overview of the work, showing the different layers of data (left panel) used in the reported studies. Gene exp = gene expression. MRI = Magnetic Resonance Imaging. NFL = Neurofilament light chain. OSR = IRCCS San Raffaele Hospital. CHUT = Centre Hospitalier Universitaire de Toulouse. KI = Karolinska Institutet.

Figure S6.1: Environmental questionnaire



E5. Per favore, indichi il peso alla Diagnosi (in kg).
Indichi 0 (zero) se non lo ricorda.

E6. Per favore, indichi la sua altezza alla Diagnosi (in cm).
Indichi 0 (zero) se non lo ricorda.

E7. Durante l'adolescenza, quanti bicchieri di bevande alcoliche assumeva mediamente in una settimana?

E8. Di che tipo? (è possibile selezionare più risposte)

Superalcolici (liquori, distillati, ...)

Vino o birra

Altro

Altro

E9. Al momento della Diagnosi di Sclerosi Multipla quanti bicchieri di bevande alcoliche assumeva mediamente in una settimana?

E10. Di che tipo? (è possibile selezionare più risposte)

Superalcolici (liquori, distillati, ...)

Vino o birra

Altro

Altro

Figure S5.1. An extract of the Environmental Questionnaire, section E (Body weight and dietary habits).